Claim 197. A DNA molecule according to Claim 164 that is single or double stranded.

Claim 198. A DNA molecule according to Claim 165 that is single or double stranded.

Remarks

Submitted herewith in support of the Application are four (4) Declarations of Dr. Richard J. Gregory, a senior inventor herein, under 37 CFR 1.131 and 1.132 identified as "A", "B", "C", and "D" that address in very great detail features that characterize the present inventions and their scope, and that further discuss the limits of the technology disclosed in the prior art references.

Support for the Amendments to the Specification and the Claims is provided as follows:

(1) With respect to the paragraph to be added to the Specification, entry of the amendment is proper since the paragraph was found in the 07/488,307 parent application and was similarly positioned in the 07/589, 295 parent application. The '295 application was not abandoned until July 13, 1993, a date which is after the July 2, 1993 filing date of the present application. Thus the applications were co-pending. Entry of this amendment was requested originally in Applicants' paper of November 22, 1994, transmitted by facsimile, which may not have reached the Examiner. A duplicate copy of that Paper is attached hereto so that the file can be made complete.

(2) The terminology used in the new claims more closely follows that found in the Specification. To facilitate consideration of the Claims, the

following Listing provides mention of where certain terms or phrases used therein can be found in the Specification:

toxic CFTR fragments — page 10, line 32; and page 11, line 3.

inviability — page 10, line 31; and page 13, line 29.

stabilized cDNA — page 5, line 7.

stably propagated — page 2, line 21 and page 4, line 14.

intervening sequence — page 4, lines 19-22; and page 7, line 30.

disrupts expression — page 4, line 20.

spliced from CFTR primary RNA — page 7, line 32-34; and page 13, line 33.

disrupts the translational reading frame — page 5, lines 8-10

disrupts the protein coding sequence — page 13, line 30.

stop codons — inherent in the sequence depicted in Figure 6

stably maintained — page 13, at line 16; page 11, at line 30; and page 8, at line 9.

origin of replication — page 4 at line 30; and page 13 at line 4.

25 copies or less — page 5 at lines 1-3.

position 936 T to C mutation — page 17 at line 20.

transformed host cell — page 8, line 10.

Preliminary Issues under Section 112

Many of the suggestions made by the Examiner in the Official Action of December 9, 1994 (hereinafter the "Official Action") have been adopted.

In response to the rejection of the claims and parallel objection to the Specification that begins on Page 2, line 10 of the Official Action and that continues to Page 3, line 10 thereof, the term "propagation" has generally been adopted for use in the new claims. Applicants note however that there is no requirement under U.S. patent law that the terms used in the claims mirror precisely the terms used in the Specification, as long as usage would be understood in the art. The term "maintained" is, in fact, used in the

Specification, as noted in the Listing presented above. New claims 173, 174, 179 and 180 refer, for example, to stable maintenance of plasmids— which is precisely a use of that term as found in the Specification.

With respect to the rejection under 35 USC § 112, first paragraph, that begins at Page 3, line 12 of the Official Action and continues to page 6, line 1 thereof, Applicants have adopted certain of the Examiner's suggestions. The subject matter of the claims that are pending herein has been limited to involve only human CFTR or human CFTR-encoding sequence. The term "sufficiently duplicative of human CFTR", and the like, has been deleted from the claims. Similarly, all of the claims have been limited to recite propagation, expression, and the like, in only one host organism , that is, E. coli. A substantial portion of this Reply is, however, directed to responding to the Examiner's rejections under 35 USC § 112, first paragraph, that concern the scope of claim coverage to be accorded in relation to "introns" and "point mutations".

With respect to the rejection under 35 USC § 112, second paragraph, that is found on page 6, lines 4-11 of the Official Action, it is believed that Applicants cancellation herewith of the involved claims makes the rejection moot. Applicants however acknowledge the Examiner's remarks.

The Section 102(a) Rejection

The Examiner has rejected previously pending Claims 139, 140, 146, and 150 under 35 USC § 102(a) as being anticipated by J. Riordan et al.,"Identification of the Cystic Fibrosis Gene: Cloning and Characterization of Complementary DNA", Science, 245, 1989, pp. 1066-1073 (hereinafter "Riordan et al. 1989"). Since the rejection is likely to be applied to the newly submitted claims it is addressed. The Riordan et al. 1989 reference discloses the existence of natural introns of the CFTR gene, and the Examiner wishes to shift to Applicants the burden of

showing why "endogenous introns would not be expected to prevent functional CFTR expression in bacteria" (see the Official Action at page 6, line 14 to page 7, line 2).

It is possible that certain natural introns, appropriately positioned, would indeed stabilize a full length CFTR-encoding DNA, or cDNA, to propagation in E. coli. This could be so if, for example, the placement of such an intron into a CFTR-encoding DNA leads to satisfaction of the functional criteria first disclosed in Applicants Specification, as set out also in Applicants' newly presented Claim 177, which criteria are further discussed in the Declaration "B" of Dr. Richard Gregory, an inventor herein, that is submitted herewith. However, submitted Declarations "A", "B", and "D" of Dr. Gregory discuss at length underlying facts about the Riordan et al. 1989 disclosure that remove it as a reference under § 102 against the claims of the present application. Accordingly, this rejection is traversed in its entirety and no burdern may be shifted to Applicants herein to compare or demonstrate any properties of the introns described by Riordan et al. 1989.

Simply stated, proper application of the novelty standard under § 102 requires that there be identity between the subject matter disclosed in the prior art reference and the subject matter of an Applicant's later application (as claimed therein), before there can be an anticipation. The present Applicants' have not claimed any introns *per se* as compositions, nor have they claimed any intron-containing DNA molecules that encode only a fragment of full length CFTR. Instead, Applicants relevant claims are, and at all times have been, directed to DNA molecules that encode full length CFTR, a composition which Riordan et al. 1989 neither disclose, nor enable production of.

Riordan et al. 1989 as published on **September 8, 1989** contains no written description whatever concerning construction of a full length CFTR-encoding DNA or cDNA, nor any disclosure concerning how to stabilize the CFTR-encoding DNA for propagation in E. coli. Although the disclosure in Document (B) contains the disclosure (at page 1072, lines 26-31 thereof)

> "With the CF gene (that is, the cDNA) now isolated, it should be possible to elucidate the control of ion transport pathways in epithelial cells in general. Knowledge gained from the study of the CF gene product (CFTR), both the normal and mutant forms, will provide a molecular basis for the development of improved means of treatment of disease."

in fact, at that time, no full length CFTR-encoding cDNA or DNA had been constructed, and the production of such stable constructs was not enabled thereby. The above quote is readily misunderstood.

There are available, however, other publications by authors/inventors having commonality of authorship/inventorship with the authors of Riordan et al. 1989, and published after that document, that provide clear "written description" of difficulties and setbacks that those authors faced in their attempts to isolate or assemble a full-length CFTR-encoding DNA and, in particular, one that was stabilized for propagation in E. coli. These experimental difficulties were so significant, and were for so long not overcome, that one is compelled to conclude that the Riordan et al. 1989 disclosure fails to enable production of a DNA or cDNA that encodes full-length CFTR. Indeed, as shown below, persons who are co-authors of the Riordan et al. 1989 reference concede this very fact. Publications of this type are discussed directly below.

(1) An early indication from authors having commonality with the authors of Riordan et al. 1989 that assembly of a full length CFTR-encoding construct might

pose unanticipated difficulties can be found in Document (D) — as it is identified in the **Exhibit Book** submitted herewith — in the paragraph thereof (referring to isolation of stable clones) that bridges pages 31-32.

> "Since the number of recombinant cDNA clones for the CF gene detected in the library screening was much less than would have been expected from the abundance of transcript estimated from the hybridization experiments, it seemed probable that the clones that contained aberrant structures were preferentially retained while the proper clones were lost during propagation. Consistent with this interpretation, poor growth was observed for the majority of the recombinant clones isolated in this study, regardless of the vector used." (emphasis added)

(2) Document (H) — again as it is identified in the Exhibit Book — is a single page abstract published in **October of 1990,** that recites in nearly each and every sentence thereof a series of experimental problems that presumably began immediately after the publication of Riordan et al. 1989. Notable of statements included therein are:

> (a) ... "conventional approaches to cloning are unlikely to work."
>
> (b) "Numerous other strategies have also been attempted without success."
>
> (c) "Reducing the quantity of product by switching from a high copy number vector (Bluescript) to a lower copy number vector (about 20 copies per cell) of the pBR322 type yields rearranged clones also.", and
>
> (d) ... "the resultant 'normal' clones were also rearranged, whereas the deleted clones were stable. The type and spacing of these clone-stabilizing mutations suggest that at least a portion of CFTR is expressed in the bacterial host and that the unaltered product is toxic to the bacterium."

With respect to item (c) listed directly above, Document (H) specifically **teaches away** from the utility of low copy number methodology which, by that time, had **already been disclosed** in the earliest parent of the present application

(United States Serial No. 07/488,307 filed March 5, 1990), along with two other methods, as being effective to stabilize the CFTR DNA for cloning in bacterial cells. Simply stated, the authors of Document H (Drumm et al. who share commonality with the authors/inventors of Riordan et al. 1989 — and with those who discovered the CFTR gene as a set of pieces) literally **walked right up to the underlying answer, could not recognize it,** and **walked away** from it never having understood that truly low copy number would work. This underscores the fact that those authors/inventors did not believe that a cryptic promoter was responsible for the difficulties encountered in cloning the CFTR cDNA — a conclusion that turned out to be clearly incorrect.

(3) Finally, and most particularly, attention should be directed to Document (J) as it is identiifed in the Exhibit Book, where at page 7 thereof, under the heading "Construction of a full-length CFTR cDNA", Dr. Lap-Chee Tsui explains **in 1991** that

> "The reconstruction of a full length cDNA from the previously isolated overlapping fragments [citing Riordan et al. 1989] has been hampered, however, by the instability of a DNA sequence within the coding region [citing Document (I) published in **September of 1990** for disclosure by those authors of the existence of the problem]. This technical difficulty was eventually overcome by the use of a low copy number bacterial plasmid [citing the work of Richard J. Gregory et al.] and ....." (emphasis added, citations therein again refer to Exhibit Book)

In summary, it is plainly apparent to those skilled in the art that Dr. Tsui, a co-inventor/co-author of Riordan et al. 1989 has all but conceded that he/they did not enable provision of a full length CFTR-encoding DNA that was stable to propagation in bacteria in Riordan et al. 1989, and that enablement of this was not disclosed until September of 1990.

[It also follows directly that any previous rejection of Applicants'claims directed to low copy number technology (now pending as claims 174 and 180) based on a

reference having such commonality of authorship/inventorship (see, for example, the rejection over Collins et al., U.S. Patent 5,240,846, as mentioned in Paper No. 24, Examiner Interview Summary Record of November 23, 1994) cannot have been maintained.]

Accordingly, and as further elaborated in paragraph 12 of Dr. Gregory's Declaration "D" , the earliest-in-time document that discloses construction of an intact full length CFTR- encoding DNA is believed to be present Applicants' United States Application Serial No. 07/488,307, filed **March 5, 1990**, which is the earliest parent of the present application. As reported therein, production of such DNA was accomplished by maintaining the DNA at low copy number in host E. coli (Examples 2 and 3 thereof), or by insertion into the CFTR-encoding cDNA of an intron that prevented expression from the cryptic bacterial promoter of RNAs that direct production of toxic CFTR fragments (Example 4 thereof). Additionally, that an equally effective outcome could be achieved by introducing a point mutation into the promoter sequence itself to disrupt recognition thereof by E. coli , but while not altering the amino acid sequence of the CFTR polypeptide to be expressed therefrom in mammalian cells, was also disclosed therein (Example 6).

The present Applicants, being first-in-time, are thus entitled also to claims directed to full length CFTR-encoding DNA and even to such DNAs that lack any kind of stabilization or sequence modification to facilitate propagation in E. coli, i.e. wild type human CFTR-encoding cDNA. However, in order to facilitate examination of this application, claims directed to wild type (not stabilized) human CFTR-encoding cDNA are presently pending in a further continuation application.

Applicants are concerned additionally that the Patent Office may restate the rejection over Riordan et al 1989, which appears to involve elements of

inherent anticipation, as a rejection based on § 103, or alternatively, as based on both § 102 and § 103. Although the making of such a rejection may conceivably be permitted in some limited circumstances (Ex parte Gray, 10 USPQ 2d 1922, Board of Patent Appeals and Interferences, 1989), this is not so here where, without reference to the disclosure in the Specification of present Applicants, the practitioner of the art would have had no knowledge of the relationship between the certain type of intron or "intervening sequence" as taught by Applicants and stable propagation in bacteria. To make such a rejection is simply to use Applicants' Specification as a reference against itself. The Examiner is requested to observe that Applicants' full length encoding constructs are vastly different—as products — from the partial clones of Riordan et al. and therefore the prerequisite of "reasonable identity" that must be in place before the Examiner can require comparison of properties is not present (see Ex parte Gray at 1924 thereof, citing In re Brown , 173 USPQ 685 (1972).

It remains to address those rejections made by the Examiner under § 112, first paragraph, that relate to the scope of claim coverage that may be afforded Applicants' pioneering inventions with respect to the use of introns and point mutations that stabilize CFTR-encoding cDNA for propagation in E. coli.

The Rejections under Section 112, First Paragraph, Concerning Introns and Point Mutations.

The Examiner has rejected under 35 USC §112, first paragraph, those of Applicants' claims that are directed to the use of introns stating that the involved disclosure is enabling only for the specific intron depicted in Figure 6 of the Specification. The Examiner has also rejected those of Applicants' claims that are directed to altering the nucleotide sequence of the cryptic promoter in the CFTR-encoding DNA by one or more point mutations stating, in this regard, that the

involved disclosure is enabling only for point mutations T748Cand A774G, as disclosed in Example 5 of the present Specification. There are additional facts that need to be called to the Examiner's attention in regard to both of these inventions, and which it is asserted, compel that the rejections be withdrawn.

The Examiner is respectfully requested to review the attached Declarations carefully as they set out a chronology of events that will greatly clarify the state of the art, the nature of Applicants' inventions, and the scope of enablement provided by the present patent application, and the parent applications thereof.

Declaration "A" is directed primarily to the state of the art at the time Applicants' inventions were made, and also establishes facts concerning the novelty of low copy number methodology.

Declaration "B" is directed primarily to those of Applicants' inventions that involve use of intervening sequences ("introns") to stabilize CFTR-encoding DNA for propagation in E. coli.

Declaration "C"is directed primarily to those of Applicants' inventions that involve using point mutations to alter the cryptic promoter in the CFTR-encoding nucleotide seqeunce such that it does not function in E. coli , but wherein the amino acid sequence encoded thereby(for later expression in eucaryotic cells) is unchanged.

Declaration "D" is also directed to Applicants' inventions that involve point mutations, and as described below, is used to swear behind certain events.

Point Mutations

By reviewing Declaration "C" and Example 5 of the present application, (see also Example 6 of the parent 07/488,307 application), the Examiner can determine that Applicants disclosedthat a nucleotide subsequence within the

CFTR-encoding cDNA, at about positions 748 to 778 thereof, was a potential promoter . Based upon their results with introns and low copy number mothodology, Applicants had already determined that a promoter was positioned in the CRTR-encoding cDNA upstream from position 1716/1717 (see Declaration "C" at paragraphs 6 to 9).

Following the methodology described in Example 5 of the application, and in routine course, the actual location of an active promoter was correctly mapped to positions 908-936 in the CFTR-encoding cDNA. This straightforward procedure is discussed at length by Dr. Gregory in his Declarations "C" and "D" attached hereto. Applicants emphasize that once existence of the promoter was disclosed by the present Applicants', isolating it was but routine.

As set out in those Declarations and following art-recognized methodology, the T936C mutation was fully reduced to practice by Gregory et al. prior to July 27, 1990, the date that their publication describing these results "Expression and Characterization of the Cystic Fibrosis Transmembrane Conductance Regulator", Nature , 347, issue of September 27, 1990, pp. 382-386, was received by that journal for publication. The invention was then disclosed in a patent application (see Example 7 of Serial No. 07/613,592, filed November 15, 1990, at page 17, line 20 thereof).

This filing date of this application is after the September 21, 1990 publication date of Document (I) as listed in the Exhibit Book, which also discloses this mutation. The filing date is also after the September 18,1990 date of the 07/584,275 application from which U.S. Patent No. 5,240,846 to Collins et al. (Document F in the Exhibit Book) issued. The '846 patent also discloses this particular mutation.

However, it is a straightforward matter for the present Applicants to swear behind both the September 18 and September 21, 1990 dates under 37 CFR

1.131, and Declarations "C" and "D" attached hereto accomplish this. Applicants note that the inventions claimed in the Collins et al. patent are patentably distinct from those that are being claimed herein so that use of the Rule 131 procedure is proper..

Applicants respectfully disagree with the Examiner's position that the scope of claim coverage to be accorded this aspect of the invention should be no broader than that encompassed by the particular species of point mutation that have been disclosed. Applicants identified a particular RNA polymerase promoter within the CFTR-encoding cDNA that conforms well to the E. coli consensus sequence. Once the existence of the promoter became known and its nucleotide sequence was determined, it became but trivial for those skilled in the art to inspect that sequence with the idea of exploiting degeneracy in the DNA code to select further suitable point mutations. No doubt, the Examiner would conclude - if Applicants had failed to claim their invention generically-that publication of their disclosure would render all such further mutations obvious. Applicants submit that the process of identifying such further suitable mutations within the already-identified promoter is trivial, and involves an amount and kind of experimentation that is well within the standards that define enablement (Ex parte Forman et al., 230 USPQ 546, Board of Patent Appeals and Interferences, 1986, and In re Wands, 8 USPQ 2d 1400, CAFC, 1988).

Intervening Sequences (introns)

For reasons similar to those provided directly above, Applicants' claims directed to use of intervening sequences (introns) to stabilize propagation in E. coli of CFTR-encoding DNAs, and in particular cDNAs, are broadly allowable. The Examiner's attention is directed respectfully to Atlas Powder Co. v. E.I.

DuPont Nemours, 750 F.2d 1569, 224 USPQ 409, (Fed Cir. 1984) at 1576 wherein the Court determined:

> We agree with the district court's conclusion on enablement. Even if some of the claimed combinations were inoperative, the claims are not necessarily invalid. "It is not a function of the claims to specifically exclude ... possible inoperative substances...(citations omitted).

Applicants' Specification provides reasonable guidance as to the factors that guide one to the selection of appropriate intervening sequences and to the positioning thereof. These factors were discussed at length in Declaration "B" of Dr. Gregory. Indeed, once Applicants determined that an active cryptic promoter was positioned at nucleotide positions 908 to 936 in the CFTR-encoding cDNA, placement of such introns, for example, at natural exon/intron boundaries close thereto is immediately suggested.

Siince use of point mutations and intervening sequences provide the only two ways that, practically speaking, could be used to achieve the benefits provided according to the practice of the invention, presentation by Applicants' of broad claims is appropriate.

Applicants reiterate the facts set out in the Declarations provided herewith which establish that no publication or patent application of another person has an effective date with respect to any of the above-described inventions that is prior to September of 1990. This is particularly important with respect to the following references (using the identifying letters of the Exhibit Book).

(C) L. Tsui et al., "Cystic Fibrosis Gene", International Patent Application Number PCT/CA90/00267 published on March 7, 1991, bearing Publication Number WO 91/002796, and claiming the priority of United States patent applications 07/396,894, 07/399,945, and 07/401,609 filed, respectively, on August 22, 24, and 31, 1989;

(E) F.S. Collins et al.,"Gene Therapy for Cystic Fibrosis", International Patent Application Number PCT/US91/06660 published on April 2, 1992, bearing Publication Number WO 92/05273, and claiming the priority of United States patent application 07/584,275 filed on September 18, 1990; and

(F) United States Patent 5,240,846 to Collins et al. issued August 31, 1993 from Application No. 07/584,275 filed September 18, 1990.

On the face of the Collins et al. patent it is stated that each of the U.S. patent applications whose priority was claimed in either of Documents (E) or (F) above was to be considered a parent appplication of the patent. However since more than a year elapsed between the filing of the priority applications that support Document (E) and the priority application that supports Document (F), international priority to all such applications could not be claimed in one PCT application. The resultant bifurcation into two PCT applications, whose disclosures can be compared, provides to the Examiner a simple tool whereby to confirm that disclosure in the Collins et al. patent, which might otherwise appear to be a reference against the present Applicants' inventions, has instead an effective date only of September 18, 1990, as reflected plainly by the text of Document (E).

Should the Examiner consider presenting art further art rejections based upon these references, or other references that are related thereto, the burden is shifted to the Patent Office to evalaute in each such reference the true dates to which each disclosure therein is entitled. In this regard, the Examiner may wish to consult

Application of Rinehart 531 F.2d 1048, at 1052 (CCPA, 1976) wherein that Court stated:

"When *prima facie* obviousness is established and evidence is submitted in rebuttal, the decision-maker must start over. Though the burden of going forward to rebut the *prima facie* case remains with the applicant, the question of whether that burden has been successfully carried requires that the entire path to decision be retraced."

Request for an Extension of Time

Pursuant to 37 CFR §§ 1.136(a), Applicants requested on June 6, 1995 an Extension of Time for three (3) months from March 9, 1995 until June 9, 1995 so that a continuation application (itself filed June 6, 1995) of this application could be filed. A duplicate of that Request marked "COPY" is attached hereto. The Patent Office is hereby authorized to charge any other fee or fee amount that it determines may still be necessary to secure the filing of this Amendment to Deposit Account **07-1074**.

Additionally, **no fees for the new claims** are believed to be due since both the number of independent claims and the number of total claims added hereinby are less that the numbers thereof that have been canceled herewith or canceled previously. However, should the Patent Office determine that claim fees are due, then such fee may be charged to Deposit Account **07-1074**.

Conclusion

Applicants do not waive their right to file a divisional or other continuation application directed to the subject matter of any claim by virtue of its having been canceled, amended, or withdrawn from consideration.

Applicants firmly believe that the application is fully in condition for allowance, and that all pertinent issues have been addressed. It is respectfully requested that the claims now be passed to allowance. The Examiner is invited to phone the undersigned or Mark A. Hofer at (508) 872-8400 to discuss any matters that she believes require further attention. An early and favorable action is respectfully requested.

Respectfully submitted,

GENZYME CORPORATION

Dated: 6/9/95

By: William G. Gosz

William G. Gosz, Esquire
Registration No. 27,787
Attorney for Applicants

One Mountain Road
Framingham, MA 01701- 9322
(508) 872-8400 x 2558

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants: Gregory et al. Examiner: Dr. K. C. Carlson

Serial No: 08/087,132 Art Unit: 1812

Filed: July 2, 1993 Docket: NZI-012CN
IG4-9.2(FWC)

For: NEW DIAGNOSTIC AND TREATMENT METHODS INVOLVING THE CYSTIC FIBROSIS TRANSMEMBRANE REGULATOR

The Honorable Commissioner
of Patents and Trademarks
Washington, DC 20231

Declarations "A" and "B" and "C" and "D" of Dr. Richard J. Gregory

SIR:

Attached hereto are the original Declarations "A" and "B" and "C" and "D" of Dr. Richard J. Gregory, an inventor herein.

Respectfully submitted,
GENZYME CORPORATION

Dated: June 9, 1995

By: William G. Gosz
William G. Gosz
Registration No. 27,787
Attorney for Applicants

One Mountain Road
Framingham, MA 01701-9322
(508) 872-8400

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants:	Gregory et. al.	Examiner:	Dr. K. C. Carlson
Serial No.:	08/087,132	Art Unit:	1812
Filed:	July 2, 1993	Docket:	NZI-012CN IG4-9.2(FWC)

For: NEW DIAGNOSTIC AND TREATMENT METHODS INVOLVING THE CYSTIC FIBROSIS TRANSMEMBRANE REGULATOR

Honorable Commissioner of
Patents and Trademarks
Washington, DC 20231

Declaration "A" of Dr. Richard J. Gregory under 37 CFR 1.132

SIR:

Dr. Richard J. Gregory declares as follows:

1. I am a co-inventor of the present U.S. Application Serial No. 08/087,132 and of the subject matter described therein, and a co-inventor also of all of the parent applications thereof — No. 07/613,592, filed on November 15, 1990, No. 07/589,295 filed September 27, 1990, and No. 07/488,307 filed March 5, 1990.

2. I received my doctoral degree from the University of Massachusetts in 1986 and was a Principal Scientist at Genzyme Corporation in Framingham, MA during the time that the inventions described in the above-identified patent applications were made.

3. I have read and am familiar with our above-identified patent application, and am familiar also with the content of the following Documents that are of record herein:

(A) J. M. Rommens et al., "Identification of the Cystic Fibrosis Gene: Chromosome Walking and Jumping", Science, 245, 1989, pp. 1059-1065.

(B) J. Riordan et al., "Identification of the Cystic Fibrosis Gene: Cloning and Characterization of Complementary DNA", Science, 245, 1989, pp. 1066-1073.

— signature copy—

(C) L. Tsui et al., "Cystic Fibrosis Gene", International Patent Application Number PCT/CA90/00267 published on March 7, 1991, bearing Publication Number WO 91/002796, and claiming the priority of United States patent applications 07/396,894, 07/399,945, and 07/401,609 filed, respectively, on August 22, 24, and 31, 1989.

(D) L. Tsui. et al. "Introns and Exons of the Cystic Fibrosis Gene and Mutations at Various Positions of the Gene", International Patent Application Number PCT/CA91/00009 published on July 25, 1991, bearing Publication Number WO 91/10734, and claiming the priority of Canadian national applications 2,007,699, 2,011,253, and 2,020,817 filed, respectively, on January 12, March 1, and July 10, 1990.

(E) F.S. Collins et al.,"Gene Therapy for Cystic Fibrosis", International Patent Application Number PCT/US91/06660 published on April 2, 1992, bearing Publication Number WO 92/05273, and claiming the priority of United States patent application 07/584,275 filed on September 18, 1990.

(F) United States Patent 5,240,846 to Collins et al. issued August 31, 1993 from Application No. 07/584,275 filed September 18, 1990

(G) L. Tsui et al., "Stable Propagation of Modified Full Length Cystic Fibrosis Transmembrane Conductance Regulator Protein cDNA in Heterologous Systems", International Patent Application Number PCT/CA91/00341 published on April 2, 1992, bearing Publication Number WO 92/05252, and claiming the priority of Great Britain national application 9020632.7 filed September 21, 1990.

(H) M. Drumm et al., "The Full Length CFTR cDNA is Toxic in Bacteria", Pediatric Pulmonology, Supplement 5 (Abstracts), October 1990, Abstract No. 8, page 189.

(I) M.L. Drumm et al., "Correction of the Cystic Fibrosis Defect in vitro by Retrovirus-Mediated Gene Transfer", Cell, 62, September 21, 1990, pp. 1227-1233.

(J) L. Tsui. "Probing the Basic Defect in Cystic Fibrosis", Current Opinion in Genetics and Development, 1, 1991, pp. 4-10.

4. My co-inventors and I commenced assembly of a full length CFTR-encoding deoxyribonucleotide sequence when it became apparent that, for unknown reasons, the authors of the above-mentioned Document (B) were unable to announce construction of a full-length CFTR-encoding DNA.

5. Accordingly, following the publication date of Document (B), my co-inventors and I obtained partial cDNA clones T11, T16-1, T16-4.5, and C1-1/5 from the American Type Culture Collection Rockville, MD, as deposited therein by Riordan et al. 1989 (see Document B).

6. Document (B) as published on **September 8, 1989** contains no written description whatever concerning construction of a full length CFTR-encoding DNA or cDNA, nor any disclosure concerning how to stabilize the CFTR-encoding DNA for propagation in E. coli. Although the disclosure in Document (B) contains the disclosure (at page 1072, lines 26-31 thereof)

> "With the CF gene (that is, the cDNA) now isolated, it should be possible to elucidate the control of ion transport pathways in epithelial cells in general. Knowledge gained from the study of the CF gene product (CFTR), both the normal and mutant forms, will provide a molecular basis for the development of improved means of treatment of disease."

in fact, at that time, no full length CFTR-encoding cDNA or DNA had been constructed, and the production of such stable constructs was not enabled thereby.

7. The following of the above-listed Documents (which in all cases have commonality of authorship/inventorship with the authors of Document B) are among those that — instead — provide clear "written description" of difficulties and setbacks that those authors faced in their attempts to isolate or assemble a full-length CFTR-encoding DNA and, in particular, one that was stabilized for propagation in E. coli .

Document(D) which incorporates several priority filing dates up to and including a filing date in **July of 1990** discloses at pages 31-32 thereof:

> "Since the number of recombinant cDNA clones for the CF gene detected in the library screening was much less than would have been expected from the abundance of transcript estimated from the hybridization experiments, it seemed probable that the clones that contained aberrant structures were preferentially retained while the proper clones were lost during propagation. Consistent with this interpretation, poor growth was observed for the majority of the recombinant clones isolated in this study, regardless of the vector used." (emphasis added)

Document (H), a single page abstract published in **October of 1990,** recites in nearly each and every sentence thereof a series of experimental problems that presumably began immediately after the publication of Document (B). Notable of statements included therein are:

> (a) ... "conventional approaches to cloning are unlikely to work."
>
> (b) "Numerous other strategies have also been attempted without success."
>
> (c) "Reducing the quantity of product by switching from a high copy number vector (Bluescript) to a lower copy number vector (about 20 copies per cell) of the pBR322 type yields rearranged clones also.", and
>
> (d) ... "the resultant 'normal' clones were also rearranged, whereas the deleted clones were stable. The type and spacing of these clone-stabilizing mutations suggest that at least a portion of CFTR is expressed in the bacterial host and that the unaltered product is toxic to the bacterium."

With respect to item (c) listed directly above, Document (H) specifically **teaches away** from the utility of low copy number methodology which, by that time, had **already been disclosed** in our earliest United States priority document (Serial No. 07/488,307 filed March 5, 1990), along with two other methods, as being effective to stabilize the CFTR DNA for cloning in bacterial cells. Simply stated, the authors of Document H (Drumm et al. who share commonality with the authors/inventors of the prior art references applicable herein — and with those who discovered the CFTR gene as a set of pieces) literally **walked right up to the underlying answer, could not recognize it,** and **walked away** from it never having understood that truly low copy number would work. This underscores the fact that those authors/inventors did not believe that a cryptic

promoter was responsible for the difficulties encountered in cloning the CFTR cDNA — a conclusion that turned out to be clearly incorrect.

As mentioned above, my co-inventors and I had on March 5, 1990 already disclosed that **truly low copy** number **was** effective. That description (see Examples 2 and 3 of Applicants' 07/488,307 priority document, or Examples 1 and 2 of the present application) provides what we believe is the first successful assembly of a full length CFTR-encoding DNA. Example 4 (intron insertion) and Example 6 (point mutation within the promoter) of that priority application disclose stabilizing the CFTR-encoding cDNA for propagation in E. coli. (see corresponding Examples 3 and 5 of the present application).

Finally, and most particularly, attention should be directed to Document (J), where at page 7 thereof, under the heading "Construction of a full-length CFTR cDNA", Dr. Lap-Chee Tsui explains **in 1991** that

> "The reconstruction of a full length cDNA from the previously isolated overlapping fragments [citing Document "B" above] has been hampered, however, by the instability of a DNA sequence within the coding region [citing Document (I) above published in **September of 1990** for disclosure by those authors of the existence of the problem]. This technical difficulty was eventually overcome by the use of a low copy number bacterial plasmid [citing the work of Richard J. Gregory et al.] and ....." (emphasis added)

In summary, it is plainly apparent to those skilled in the art that Dr. Tsui, a co-inventor/co-author of nearly every Document enumerated above has all but conceded that he/they did not enable provision of a full length CFTR-encoding DNA that was stable to propagation in bacteria prior to **September of 1990**.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of United States Code, and that the making of any such willful false statements may jeopardize the validity of this application, of related applications, and of any patent issued thereon.

Date: 6/8/95

Declarant: Richard J. Gregory
Richard J. Gregory, Ph.D.

— signature copy—

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants:	Gregory et. al.	Examiner:	Dr. K. C. Carlson
Serial No.:	08/087,132	Art Unit:	1812
Filed:	July 2, 1993	Docket:	NZI-012CN IG4-9.2(FWC)

For: NEW DIAGNOSTIC AND TREATMENT METHODS INVOLVING THE CYSTIC FIBROSIS TRANSMEMBRANE REGULATOR

Honorable Commissioner of
Patents and Trademarks
Washington, DC 20231

Declaration "B" of Dr. Richard J. Gregory under 37 CFR 1.132

SIR:

Dr. Richard J. Gregory declares as follows:

1. I am a co-inventor of the present U.S. Application Serial No. 08/087,132 and of the subject matter described therein, and a co-inventor also of all of the parent applications thereof — No. 07/613,592, filed on November 15, 1990, No. 07/589,295 filed September 27, 1990, and No. 07/488,307 filed March 5, 1990.

2. I received my doctoral degree from the University of Massachusetts in 1986 and was a Principal Scientist at Genzyme Corporation in Framingham, MA during the time that the inventions described in the above-identified patent applications were made.

3. I have read and am familiar with our above-identified patent application, the content of the Official Action of December 9, 1994 that is pending herein, and the reference **Riordan et al.** (J. Riordan et al., "Identification of the Cystic Fibrosis Gene: Cloning and Characterization of Complementary DNA", Science, 245, 1989, pp. 1066-1073) that is cited in the Official Action against certain claims of our application.

I. The Scope of the Riordan et al. Disclosure

4. As explained in my Declaration "A" and Declaration "D"submitted herewith, Riordan et al. (see also its companion paper, J. M. Rommens et al., "Identification of the Cystic Fibrosis Gene: Chromosome Walking and Jumping", Science, 245, 1989, pp. 1059-1065) describe the isolation of the gene for CFTR (and

the corresponding cDNA) only as a series of overlapping partial clones, but with no clone encoding full length CFTR having been isolated or constructed. My Declarations "A" and "D" describe — in considerable detail — the difficulties, failures, and setbacks encountered by authors/inventors having commonality with Riordan et al. with respect to their subsequent attempts to asssemble a full length DNA or cDNA coding for CFTR, and in particular, such a DNA or cDNA stabilized for propagation in E. coli.

5. I have noted the Examiner's remarks in the Office Action of December 9, 1994 that reflect her suggestion that natural introns, such as those disclosed by Riordan et al., may be useful to stabilize CFTR-encoding cDNA for propagation in E. coli, and that this aspect of the Riordan et al. disclosure anticipates certain of our inventions.

6. Newly pending Claim 168 defines certain inventions that relate to CFTR-encoding DNAs that contain intervening sequences (see the definition of "intron" on page 7 of the Specification). The defined elements of these inventions include inter alia that the involved DNAs be purified and isolated, that they encode full length CFTR, are stabilized to propagtion in E. coli owing to placement therein of an intervening sequence capable of being spliced from CFTR primary RNA transcript when expressed in a eucaryotic cell, and wherein said intervening sequence is placed downstream from the cryptic promoter in the encoding DNA, and contains one or more stop codons or disrupts the CFTR translational reading frame. Riordan et al. anticipates none of these.

7. Additionally, Riordan et al. neither disclose — nor even remotely suggest— use of any kind of intervening sequence to stabilize DNA for propagation in bacteria. In fact, the first reference that teaches such use of introns is U. S. Serial No. 07/488,307, which is a parent application of the present application.

II. Intervening Sequences ("introns") of the Present Invention

8. Example 4 of our parent 07/488,307 Specification (which is Example 3 of the present Specification) describes the construction of a full-length CFTR-encoding cDNA that contains an intervening sequence or "intron" that was placed bewteeen nucleotide positons 1716 and 1717 of the cDNA.

— signature copy —

9. That one or more E. coli RNA polymerase promoters were indeed present within the nucleotide sequence of the CFTR cDNA was reasonably predicted from experiments (Examples 2 and 3 of the parent 07/488,307 Specification which are Examples 1 and 2 of the present Specification) whereinby we established that the CFTR-encoding cDNA could indeed be propagated in E. coli — if care was taken to place it in a plasmid maintained therein at very low copy number.

10. As described more fully in Declaration "C" attached hereto, I performed an analysis of the nucleotide sequence of the CFTR-encoding cDNA prior to the filing of our 07/488,307 parent application. That analysis indicated that one or more nucleotide subsequences in the cDNA, and located upstream from position 1716 of the sequence, have strong sequence homology with the consensus sequence for an E. coli RNA polymerase promoter.

11. We were able to confirm that placement of the intron described in Example 3 of the present Specification (see also Figure 6 thereof) between nucleotide positions 1716/1717 of the CFTR-encoding cDNA conferred advantageous growth characteristics on host E. coli cells.

This determination was based, in part, upon the following factors:

(A) The intron was positioned downstream from two specific sites (see Exhibit 1 to Declaration "C") where the promoter was expected to be located, but upstream from the nucleotide subsequence(s) of the CFTR cDNA that code for most of the presumptive transmembrane domains of CFTR (Riordan et al. at 1071). Since CFTR is a transmembrane protein, we presumed that at least a portion of its substantial toxicity toward E. coli host cells harboring the cDNA was due to insertion into the bacterial cell membrane of CFTR fragments produced via expression of the CFTR-encoding cDNA resulting from these E. coli promoter sequences. We also presumed that a sufficiently truncated CFTR (i.e. a CFTR fragment) would be less likely to exhibit severe toxicity toward E. coli ;

(B) The intron disclosed in Figure 6 of the present application (and which is representative of introns useful in the practice of the invention) contains 83 bases (not a multiple of 3) so that read-through by E. coli necessarily involves a frameshift of the translational reading of the CFTR, so that the toxic peptide sequence would not be translated. Additionally, the intron of Figure 6 contains stop codons to further increase the chance that read-through will not occur; and

(C) The nucleotide 1716/1717 boundary defines a natural exon/intron boundary. The use of an intron inserted into this site is advantageous since the intron will be efficiently spliced from the CFTR primary transcript when placed in eucaryotic cells.

12. Based upon my experience given the teachings of the application, it would be routine for those skilled in the art to achieve similar results by preparing other intervening nucleotide sequences (introns), and then positioning them so that results equivalent to those shown in Example 3 of the present patent application, i.e. stable propagation of the cDNA in E.coli host cells, are achieved.

Ideally, provision of these other intron-containing cDNAs involves:

(a) placing the intron dowstream from the region of CFTR cDNA that contains a cryptic promoter;
(b) placement of the intron at a natural exon/intron boundary to facilitate splicing;
(3) constructing the intron to contain a total number of nucleotides that shifts the translational reading frame dowstream therefrom, or to contain stop codons, or both; and
(4) placing the intron sufficiently close to the cryptic promoter that the truncated CFTR- fragments that are prodcued are relatively short and unlikely to contain sufficient of the CFTR polypeptide seqeunce to exhibit toxic properties.

13. It is well within the average skill of those in our art to prepare other of such constructs and to screen them to confirm stable propagation of CFTR-encoding cDNA in E. coli host cells.

14. Additional positions within the CFTR-encoding cDNA that represent natural exon/intron boundaries where, it is expected, such intervening sequences can be placed with a high likelihood of utility include the boundary at 1001/1002 (Riordan et al. at page 1068) and others that are not far downstream from this site.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of United States Code, and that the making of any such willful false statements may jeopardize the validity of this application, of related applications, and of any patent issued thereon.

Date: 6/8/95

Declarant: Richard J. Gregory
Richard J. Gregory, Ph.D.

— signature copy —

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants:	Gregory et. al.	Examiner:	Dr. K. C. Carlson
Serial No.:	08/087,132	Art Unit:	1812
Filed:	July 2, 1993	Docket:	NZI-012CN IG4-9.2(FWC)

For: NEW DIAGNOSTIC AND TREATMENT METHODS INVOLVING THE CYSTIC FIBROSIS TRANSMEMBRANE REGULATOR

Honorable Commissioner of
Patents and Trademarks
Washington, DC 20231

Declaration "C" of Dr. Richard J. Gregory under 37 CFR 1.131 and 1.132

S I R:

Dr. Richard J. Gregory declares as follows:

1. I am a co-inventor of the present U.S. Application Serial No. 08/087,132 and of the subject matter described therein, and a co-inventor also of all of the parent applications thereof — No. 07/613,592, filed on November 15, 1990, No. 07/589,295 filed September 27, 1990, and No. 07/488,307 filed March 5, 1990.

2. I received my doctoral degree from the University of Massachusetts in 1986 and was a Principal Scientist at Genzyme Corporation in Framingham, MA during the time that the inventions described in the above-identified patent applications were made.

3. I have read and am familiar with our above-identified patent application, and the content of the Official Action of December 9, 1994 that is pending herein.

4. Examples 1 and 2 of the present Specification (see also Examples 2 and 3 of the parent 07/488,307 Specification) describe experiments that were conducted in order to provide CFTR-encoding cDNA in a form that was stable to propagation in E. coli. As described in those examples, placing the cDNA in a plasmid that is maintained in E. coli at very low copy number permitted stable propagation. The success of these experiments provided important evidence that a cryptic RNA polymerase promoter was present within the CFTR-encoding cDNA. Although the potential presence in the cDNA of other regulatory elements might have been responsible for the death of bacterial clones in which the cDNA was maintained at higher copy number, we determined that the most

logical explanation was the presence in the cDNA of a cryptic RNA polymerase promoter.

5. Exhibit 1 attached to this Declaration is a photocopy of a page from my laboratory notebooks (as maintained by the Genzyme Corporation) that was properly dated by me, and properly witnessed by another, all on a date that is prior to the filing of our first priority application herein, U.S. Serial No. 07/488,307 filed March 5, 1990. The page number assigned to that notebook page, and its date, have been removed as is permitted by PTO practice. It is hereby certified that all the markings and writings on this page that highlight certain nucleotide positions **are original**.

6. As shown by Exhibit 1, contained within the region of the cDNA that originates in exon 6 of the CFTR gene are located two nucleotide subsequences (the first between about positions 748 and 778 and a second between about positions 908 and 936) having strong homology to the recognized consensus E. coli RNA promoter sequence. The Reznikoff and McClure article that was cited in Example 6 of our 07/488,307 application (see corresponding Example 5 of the present application) describes and depicts this consensus sequence.

7. **Circled** in Exhibit 1 are two potential translation initiator methionines (located at about position 980) that are appropriately downstream from the second of the two potential promoters. Directly upstream from these methionines (positions 963-967) there is also a **bracketed** sequence that shows homology with that of a ribosome binding site (a Shine-Dalgarno sequence). Exhibit I also depicts an appropriate methionine and potential Shine-Dalgarno sequence downstream from the other potential promoter.

8. Once we had determined that promoter activity was very likely found within exon 6 of the CFTR-encoding sequence, we set about demonstrating that the CFTR cDNA could be stabilized against the toxic effects of the promoter for propagation in E. coli. We were soon able to disrupt production of toxic CFTR-derived polypeptides from the promoter by placing an intervening sequence (an intron) downstream from, but still reasonably close to, the promoter-affected region of exon 6 of a CFTR-encoding cDNA. The intron was placed at the natural exon/intron boundary represented by nucleotide positions 1716/1717, which also provides a convenient cloning site. As confirmed in those experiments (page 15, lines 8-11 of the 07/488,307 parent application, and page 15, lines 1-3 of the present application) growth characteristics were improved for host E. coli cells containing the new stabilized cDNA.

— signature copy —

9. Based on generally available information and following the teachings of the application, it would be well within the skill of those in our art to determine the exact nucleotide sequences within exon 6 that act as RNA polymerase promoters therein, and to then make one or more suitable point mutations in the promoter that would prevent its recognition by E. coli — but not alter the amino acids encoded therefrom . Such an encoding cDNA would permit production of the authentic human polypeptide sequence when placed in a eucaryotic cell. The straightforward protocol whereinby we identified the precise location of the RNA polymerase promoter in the CFTR-encoding cDNA is described in detail in our publication Gregory et al., "Expression and Characterization of the Cystic Fibrosis Transmembrane Conductance Regulator", Nature , 347, 1990, pp. 382-386, which is appended to my Declaration "D", and discussed therein in detail.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of United States Code, and that the making of any such willful false statements may jeopardize the validity of this application, of related applications, and of any patent issued thereon.

Date: 6/8/95

Declarant: Richard J. Gregory
Richard J. Gregory, Ph.D.

— signature copy —

From Page No.

EXHIBIT 1

```
601  TTTAGTTTGATTTATAAGAAGACTTTAAAGCTGTCAAGCCGTGTTCTAGATAAAATAAGT

     I G Q L V S L L S N N L N K F D E G L A  19
661  ATTGGACAACTTGTTAGTCTCCTTTCCAACAACCTGAACAAATTTGATGAAGGACTTGCA

     L A H F V W I A P L Q V A L L M G L I W  21
721  TTGGCACATTTCGTGTGGATCGCTCCTTTGCAAGTGGCACTCCTCATGGGGCTAATCTGG

     E L L Q A S A F C G L G F L I V L A L F  23
781  GAGTTGTTACAGGCGTCTGCCTTCTGTGGACTTGGTTTCCTGATAGTCCTTGCCCTTTTT

     Q A G L G R M M M K Y R D Q R A G K I S  25
841  CAGGCTGGGCTAGGGAGAATGATGATGAAGTACAGAGATCAGAGAGCTGGGAAGATCAGT

     E R L V I T S E M I E N I Q S V K A Y C  27
901  GAAAGACTTGTGATTACCTCAGAAATGATTGAAAATATCCAATCTGTTAAGGCATACTGC

     W E E A M E K M I E N L R Q T E L K L T  29
961  TGGGAAGAAGCAATGGAAAAAATGATTGAAAACTTAAGACAAACAGAACTGAAACTGACT

     R K A A Y V R Y F N S S A F F F S G F F  31
1021 CGGAAGGCA
```

potential promoter elements

potential initiator methionines

Many of these potential regulatory signals could be eliminated without changing coding amino acid sequences.

In addition direct repeat sequences which increase instability of clones could be mutated as well.

II) a) 1 ug of T16-1 CFTR clone digested with Sal I for 1 hour (10 units)

b) DNA end filled with 1mM dNTPs and 1 unit of Klenow polymerase in 50ul volume. Ø-Ot extracted and ethanol ppted.

c) ~~DNA religated and used to transform XL1 Blue cells to Amp~~ next week

To Page No.

Witnessed & Understood by me, Date

Invented by Paul J. Durga Date

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants:	Gregory et. al.	Examiner:	Dr. K. C. Carlson
Serial No.:	08/087,132	Art Unit:	1812
Filed:	July 2, 1993	Docket:	NZI-012CN IG4-9.2(FWC)

For: NEW DIAGNOSTIC AND TREATMENT METHODS INVOLVING THE CYSTIC FIBROSIS TRANSMEMBRANE REGULATOR

Honorable Commissioner of
Patents and Trademarks
Washington, DC 20231

Declaration "D" of Dr. Richard J. Gregory under 37 CFR 1.131

SIR:

Dr. Richard J. Gregory declares as follows:

1. I am an inventor of the claims of the above-identified patent application and of the subject matter described therein.

2. I received my doctoral degree from the University of Massachusetts in 1986 and was a Principal Scientist at Genzyme Corporation in Framingham, MA during the time that the inventions described in the above-identified patent application were made.

3. I am a co-inventor of the present U.S. Application Serial No. 08/087,132 and of the inventions defined by Claims 164 to 167 that are presently pending therein, and a co-inventor also of all of the parent applications thereof — No. 07/613,592, filed on November 15, 1990, No. 07/589,295 filed September 27, 1990, and No. 07/488,307 filed March 5, 1990.

4. The inventions defined by the aforementioned Claims 164 to 167 were conceived of, and fully and actually reduced to practice, at a time prior to **July 27, 1990**, the date that the manuscript corresponding to the publication, Richard J. Gregory et al., "Expression and Characterization of the Cystic Fibrosis Transmembrane Conductance Regulator", Nature, 347, issue of September 27, 1990, pp. 382-386, was received by that journal for publication. The **July 27, 1990** receipt date and the August 31, 1990 acceptance date are clearly printed on the last page of this publication (which is already of record in the file of the present

— signature copy —

application). A reprint of this publication is marked Exhibit "1" and is attached to the present Declaration.

5. The Gregory et al. article, as received by Nature on **July 27, 1990,** provides simple and direct confirmation of the disclosure in our 07/488,307 parent application filed March 5, 1990 that the encoding DNA sequence for human CFTR contains a functional (cryptic) RNA polymerase promoter located upstream from nucleotide position 1716 of the cDNA. This confirmation was accomplished in a simple and straightforward fashion by constructing clones in which various small fragments of the CFTR cDNA were placed upstream from a promoter-less reporter gene, and then monitoring for transcription/translation from the reporter gene to confer resistance in the presence of antibiotic. As shown in Figure 1 of our Nature publication, the CFTR cDNA sequence fragment 905-950 confers substantial antibiotic resistance, and contains a subsequence thereof (nucleotides 908-936) that lines up well with the E. coli consensus promoter sequence (position 936 thereof provides the highly conserved "T") as depicted in the Reznikoff and McClure reference (copy attached as Exhibit "2" hereto) that was cited in Example 5 of the present application (Example 6 of our March 5, 1990 07/488,307 parent application) for the purposes of describing how to identify such a promoter .

6. I am familiar with the content of the following references that are of record herein and that have effective dates before **July 27, 1990**:

(A) J. M. Rommens et al., "Identification of the Cystic Fibrosis Gene: Chromosome Walking and Jumping", Science, 245, 1989, pp. 1059-1065.

(B) J. Riordan et al., "Identification of the Cystic Fibrosis Gene: Cloning and Characterization of Complementary DNA", Science, 245, 1989, pp. 1066-1073.

(C) L. Tsui et al., "Cystic Fibrosis Gene", International Patent Application Number PCT/CA90/00267 published on March 7, 1991, bearing Publication Number WO 91/002796, and claiming the priority of United States patent applications 07/396,894, 07/399,945, and 07/401,609 filed, respectively, on August 22, 24, and 31, 1989.

(D) L. Tsui. et al. "Introns and Exons of the Cystic Fibrosis Gene and Mutations at Various Positions of the Gene", International Patent Application Number PCT/CA91/00009 published on July 25, 1991, bearing Publication Number WO 91/10734, and claiming the priority of Canadian national applications 2,007,699, 2,011,253, and 2,020,817 filed, respectively, on January 12, March 1, and July 10, 1990.

7. I am also familiar with the content of the following references that are of record herein and that, as discussed below, have effective dates that are after **July 27,1990**:

(E) F.S. Collins et al.,"Gene Therapy for Cystic Fibrosis", International Patent Application Number PCT/US91/06660 published on April 2, 1992, bearing Publication Number WO 92/05273, and claiming the priority of United States patent application 07/584,275 filed on September 18, 1990.

(F) United States Patent 5,240,846 to Collins et al. issued August 31, 1993 from Application No. 07/584,275 filed September 18, 1990

(G) L. Tsui et al., "Stable Propagation of Modified Full Length Cystic Fibrosis Transmembrane Conductance Regulator Protein cDNA in Heterologous Systems", International Patent Application Number PCT/CA91/00341 published on April 2, 1992, bearing Publication Number WO 92/05252, and claiming the priority of Great Britain national application 9020632.7 filed September 21, 1990.

(H) M. Drumm et al., "The Full Length CFTR cDNA is Toxic in Bacteria", Pediatric Pulmonology , Supplement 5 (Abstracts), October 1990, Abstract No. 8, page 189.

(I) M.L. Drumm et al., "Correction of the Cystic Fibrosis Defect in vitro by Retrovirus-Mediated Gene Transfer", Cell, 62, September 21, 1990, pp. 1227-1233.

(J) L. Tsui. "Probing the Basic Defect in Cystic Fibrosis", Current Opinion in Genetics and Development , 1, 1991, pp. 4-10.

8. Careful review of references (A) through (C) demonstrates that there is not therein disclosed a complete DNA molecule that encodes for the human CFTR protein. Instead, the gene (or cDNA) is disclosed in pieces, with no teaching as to how one would assemble a genomic DNA (or a cDNA) that would encode full length CFTR.

9. The disclosure found in Document (B) appears to be misleading in that it is represented therein that the cDNA for CFTR has been isolated. In this regard, see Document (B) at page 1072, lines 26-31 thereof, wherein it is stated:

> "With the CF gene (that is, the cDNA) now isolated, it should be possible to elucidate the control of ion transport pathways in epithelial cells in general. Knowledge gained from the study of the CF gene product (CFTR), both the normal and mutant forms, will provide a molecular basis for the development of improved means of treatment of disease."

Careful review of Documents (B) and (C) indicates that the gene (or cDNA) is, in fact, only disclosed in pieces, that is, as a series of overlapping clones. Most particularly, neither reference suggests provision of a DNA that can be stably propagated in bacteria without causing inviability thereof.

10. An early indication from authors having commonality with the authors of Documents (A), (B), and (C) that assembly of a full length CFTR-encoding construct might pose unanticipated difficulties can be found in Document (D) in the paragraph (referring to isolation of stable clones) that bridges pages 31-32.

> "Since the number of recombinant cDNA clones for the CF gene detected in the library screening was much less than would have been expected from the abundance of transcript estimated from the hybridization experiments, it seemed probable that the clones that contained aberrant structures were preferentially retained <u>while the proper clones were lost during propagation</u>. Consistent with this interpretation, poor growth was observed for the majority of the recombinant clones isolated in this study, regardless of the vector used." (emphasis added)

11. An additional publication which shows that a full length CFTR-encoding DNA could not be prepared is Document (H) published in **October 1990**. Again, the authors thereof share commonality with the inventors/authors of references that have been cited against the present application. Notable statements by

those authors concerning their efforts to assemble such a full length construct include (a) "conventional approaches to cloning are unlikely to work", (b) "numerous other strategies have also been attempted without success", and (c) "reducing the quantity of product by switching from a high copy number vector (Bluescript) to a lower copy number vector (about 20 copies per cell) of the pBR322 type yields rearranged clones also".

12. The earliest-in-time document that discloses construction of an intact full length CFTR- encoding DNA is believed to be our United States Application Serial No. 07/488,307, filed **March 5, 1990**, which is the earliest parent of the present application. As reported therein, production of such DNA was accomplished by maintaining the DNA at low copy number in host E. coli (Examples 2 and 3 thereof), or by insertion into the CFTR-encoding DNA of an intron that prevented expression from the cryptic bacterial promoter of RNAs that direct production of toxic CFTR fragments (Example 4 thereof). Additionally, that an equally effective outcome could be achieved by introducing a point mutation into the promoter sequence itself to disrupt recognition thereof by E. coli , but while not altering the amino acid sequence of the CFTR polypeptide to be expressed therefrom in mammalian cells, was also disclosed therein (Example 6).

13. Document (I) contains disclosure that concerns production of full length CFTR- encoding DNA, and that concerns the use of a point mutation at position 936 in the CFTR cDNA to facilitate cloning in bacteria. Such disclosure is to be found in Document (I) at page 2 thereof. Document (I) has a publication date of **September 21, 1990** and was not even submitted to Cell for publication until on or about September 7, 1990. Additionally, Document (I) contains the acknowledgment (at page 1227 thereof) that " Early attempts to reconstitute a full-length CFTR cDNA from overlapping clones were uniformly unsuccessful."

14. Documents (E), (F) and(G) each contain similar disclosure (see, for example, (E) at page 14, lines 1-29, (F) at column 11, lines 17-56, and (G) at page 12, lines 9-30. However each of these documents is based on a patent application that was filed only in **September of 1990.**

15. Finally, and most particularly, attention should be directed to Document (J), where at page 7 thereof, under the heading "Construction of a full-length CFTR cDNA", Dr. Lap-Chee Tsui concedes **in 1991** that

> "The reconstruction of a full length cDNA from the previously isolated overlapping fragments [citing Document "B" above] has been hampered, however, by the instability of a DNA sequence within the coding region [citing Document (I) above published in **September of 1990** for disclosure by those authors of the existence of the problem]. This technical difficulty was eventually overcome by the use of a low copy number bacterial plasmid [citing the work of Richard J. Gregory et al.] and ....." (emphasis added)

In summary, it is plainly apparent to those skilled in the art that Dr. Tsui, a co-inventor/co-author of nearly every Document enumerated above has all but conceded that he/they did not enable provision of a full length CFTR-encoding DNA that was stable to propagation in bacteria prior to **September of 1990.**

16. All of the inventions disclosed in our present application and in all of the parent applications thereof were made in the United States.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of United States Code, and that the making of any such willful false statements may jeopardize the validity of this application, of related applications, and of any patent issued thereon.

Date: 6/8/95

Declarant: Richard J. Gregory
Richard J. Gregory, Ph.D.

— signature copy —

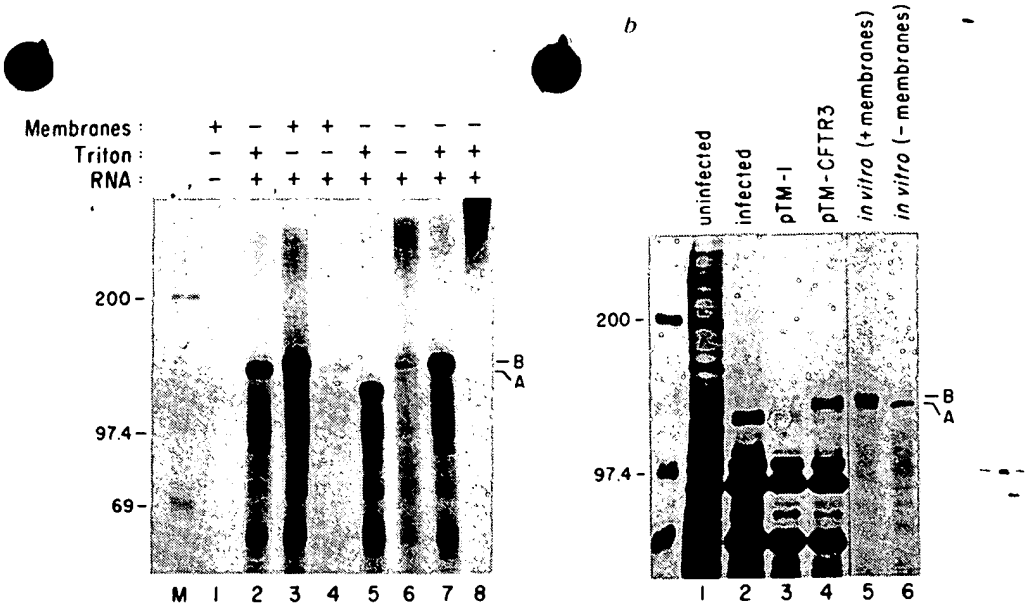


FIG. 2 *In vitro* and *in vivo* expression of the CFTR protein. *a*, CFTR cDNA transcripts were synthesized from plasmids pTM-CFTR-3 and pTM-CFTR-3ΔF508 using T7 RNA polymerase and translated in a rabbit reticulocyte lysate cell-free system. Translation products were analysed by electrophoresis on 6% SDS-polyacrylamide gels[28] and fluorography. Lane 1, control (no RNA added); lane 2, pTM-CFTR-3 transcript plus Triton X-100; lane 3, pTM-CFTR-3 transcript plus microsomal membranes; lane 4, as in lane 3 then treated with *N*-Glycanase[29]; lane 5, pTM-CFTR-3 transcript truncated after codon 1,336, plus Triton X-100; lane 6, pTM-CFTR-transcript translated without Triton X-100 or microsomal membranes; lane 7, pTM-CFTR-3ΔF508 transcript plus Triton X-100; lane 8, as in lane 2 except boiled for 3 min before electrophoresis. *b*, Expression of CFTR in transfected HeLa cells. CFTR expression was directed by a vaccinia virus/bacteriophage T7 hybrid expression system where transcription of transfected genes is catalysed by T7 RNA polymerase provided *in trans* by a recombinant vaccinia virus vector[13,17]. Cell lysates were prepared from HeLa cells labelled with [^{35}S]methionine and analysed as above. Lane 1, uninfected HeLa cells; lane 2, HeLa cells infected with vaccinia virus ; lane 3, vaccinia-infected cells transfected with the control plasmid pTM1; lane 4, vaccinia-infected cells transfected with pTM-CFTR-3; lane 5, pTM-CFTR-3 *in vitro* translation product prepared in the presence of microsomal membranes; lane 6, as in lane 5 but prepared in the presence of 0.5% Triton X-100. Protein size markers (K) shown on the left.

METHODS. *a*, Plasmids pTM-CFTR-3 and pTM-CFTR-3ΔF508 are derivatives of pSC-CFTR-2 and pSC-CFTR-2ΔF508 (Fig. 1) in which the T7 RNA promoter and EMC virus internal ribosome entry sequence from plasmid pTM-1 were placed immediately upstream of the CFTR initiation codon and the T7 terminator from pTM-1 was placed downstream of CFTR nucleotide 5,576 (refs 13, 17). For *in vitro* synthesis of full-length CFTR (1,480 amino acids), plasmids were linearized at a *Sal*I site located downstream of the CFTR sequences and transcribed with T7 RNA polymerase (Stratagene). A truncated form of CFTR was synthesized from plasmids linearized with *Tth*111I, which restricts within codon 1,337 of the CFTR sequence. One μg of each transcript was translated in a nuclease-treated rabbit reticulocyte lysate system (Promega) in a volume of 50 μl. Where indicated, 0.5% Triton X-100 or 7.2 equivalents of canine pancreatic microsomal membranes (Promega) were included in the reaction. *N*-Glycanase treatment of *in vitro* translation products was performed on 2 μl of translate (Genzyme). Reactions were diluted with an equal volume of sample buffer and 1 μl was electrophoresed, except in the case of the *N*-Glycanase reaction, in which 15 μl was analysed. Unless indicated, samples were left at room temperature for 10 min before electrophoresis and fluorography. The gel was exposed to film for 16 h. *b*, HeLa cells were plated at 50% confluency in 35-mm dishes 24 h before infection with recombinant vaccinia virus vT7-3 (ref. 13) (multiplicity of infection 10–20) and transfected with plasmids (5 μg) using 20 μg lipofectin[30] (BRL). Thirteen hours post-transfection, cells were labelled with [^{35}S]methionine (25 μCi ml^{-1}) for 1 h in hypertonic media (190 mM NaCl) before collecting in 250 μl 135 mM NaCl, 20 mM Tris-HCl (pH 7.4), 1 mM EDTA, 0.1% SDS, 1% Triton X-100, 100 μg μl^{-1} aprotinin and 17 μM PMSF. HeLa lysate (20 μl) or 1 μl of *in vitro* translate were mixed with an equal volume of sample buffer and analysed as above. To produce equivalent exposures, lanes 1–4 were exposed for 48 h and lanes 5 and 6 for 6 h.

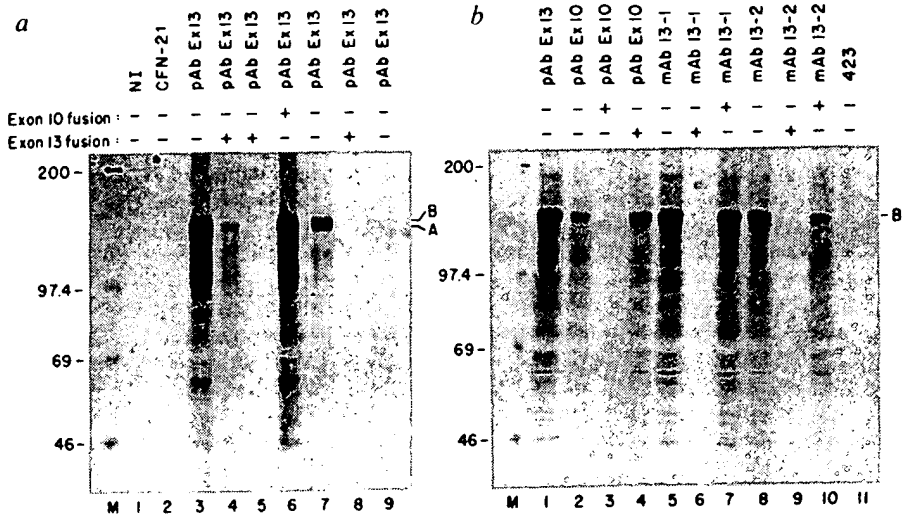


FIG. 3 Immunoprecipitation of CFTR synthesized *in vitro* in rabbit reticulocyte lysates and *in vivo* in HeLa cells. *A* Plasmid pTM-CFTR-3 was transcribed with T7 RNA polymerase and translated in a rabbit reticulocyte lysate in the presence of canine microsomal membranes and [^{35}S]methionine. Samples were immunoprecipitated with non-immune (NI) rabbit serum (lane 1), a control irrelevant polyclonal antibody (CFN-21) raised against the c-*fyn* gene product[19] (lane 2), or polyclonal antibody Ex13 (lanes 3, 4, 5 and 6). Similarly, immunoprecipitates derived from antibody Ex13 were prepared from lysates of [^{35}S]methionine-labelled HeLa cells infected with vaccinia virus and transfected with either pTM-1 (lane 9) or pTM-CFTR (lanes 7 and 8). Some reactions contained 5 μg (lane 4) or 10 μg (lanes 5 and 8) β-galactosidase–exon 13 fusion protein or 10 μg β-galactosidase–exon 10 fusion protein (lane 6). Fluorography was for 12 h. *b*, [^{35}S]methionine labelled CFTR synthesized *in vitro* and immunoprecipitated with antibody Ex13 (lane 1), Ex10 (lanes 2–4), monoclonal antibodies 13-1 (lanes 5–7) and 13-2 (lanes 8–10) raised against the β-galactosidase–exon 13 fusion protein or an irrelevant monoclonal antibody (423) raised to SV40-T (ref. 20) (lane 11). Lanes 3, 7 and 10 contained exon 10 fusion protein (10 μg) and lanes 4, 6 and 9, exon 13 fusion protein (10 μg). Fluorography was for 12 h. Preparation and purification of antibodies, and of fusion proteins were as previously published[18,19,31–33]. Protein size markers (K) shown on the left.

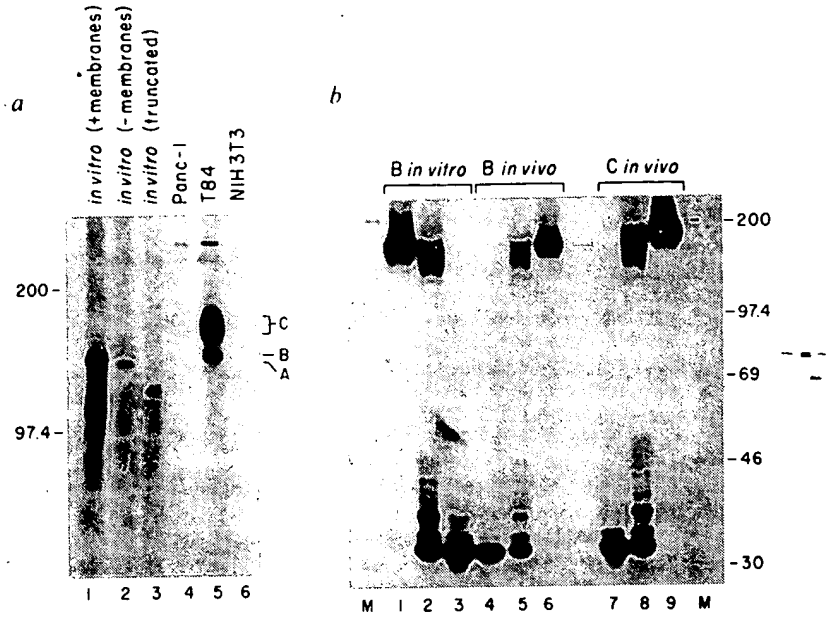


FIG. 4 *a*, Immunoprecipitation of CFTR from nonrecombinant cells. Full-length CFTR synthesized *in vitro* in the presence (lane 1) or absence (lane 2) of membranes and truncated CFTR synthesized *in vitro*, terminating at residue 1,336 (lane 3) were immunoprecipitated with antibody Ex13 and incubated with protein kinase A (20 ng) and [γ^{32}P]ATP (10 μCi) in a final volume of 50 μl in kinase buffer (50 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$ and 100 μg ml^{-1} bovine serum albumin) at 30 °C for 60 min[22]. Immunoprecipitates of equivalent amounts of extracts of Panc-1 (lane 4), T84 (lane 5) and NIH 3T3 fibroblasts (lane 6) were subjected to the same treatment. Samples were heated at 37 °C for 10 min before electrophoresis. Exposure time was 20 min at −70 °C. *b*, One-dimensional peptide analysis of CFTR synthesized *in vitro* and *in vivo*. The CFTR synthesized in reticulocyte lysate in the presence of membranes (band B) and labelled with protein kinase A and [γ^{32}P]ATP (lanes 1–3) and that immunoprecipitated from T84 lysates (lanes 4–6, Band B and lanes 7–9, Band C) were digested with increasing amounts of *S. aureus* V8 protease[19]. Proteins in lanes 2, 5 and 8 were digested with 0.017 μg μl^{-1} of *S. aureus* V8 protease, lanes 3, 4 and 7 with 0.17 μg μl of enzyme. Lanes 1, 6 and 9 were untreated samples. Exposure time was 24 h.

excised ^{32}P-labelled band B from immunoprecipitates of material synthesized in the cell-free system and prepared a partial preteolysis fingerprint using *Staphyloccus aureus* V8 protease. In parallel, we prepared similar fingerprints of bands B and C from ^{32}P-labelled immunoprecipitates of T84 cells. As seen in Fig. 4*b*, the partial proteolysis fingerprints are virtually identical, demonstrating that the proteins immunoprecipitated from T84 cells are related to CFTR synthesized *in vitro*. Similar conclusions were drawn from experiments using endoproteinases Asp-N and Lys-C (not shown). One interpretation of this result is that band C represents a fully processed version of the CFTR protein, with the addition of more carbohydrate or phosphate, or both, to the core-glycosylated CFTR. The related multiple drug resistance P-glycoprotein undergoes a similar shift in mobility on addition of *N*-linked carbohydrate[16].

These studies constitute a formal identification of the product of the CFTR open reading frame and verify several of the predictions made about the CFTR protein. First, CFTR associates with membranes and is glycosylated, properties consistent with the finding that the CF defect in Cl^- channel regulation is observed in isolated cell-free patches of membrane. Second, CFTR can be phosphorylated by cAMP-dependent protein kinase, consistent with the finding that cAMP regulates Cl^- channels in normal but not CF cell membranes. Third, CFTR is present in T84 cells but not fibroblasts, consistent with the finding that T84 cells show Cl^- channel regulation similar to that observed in normal airway epithelial cells, whereas channel regulation in fibroblasts is controversial. Furthermore, the agents described here provide materials for a wide range of studies on the CFTR and its function: for example, in the accompanying paper[25] we describe complementation of the defect in ion transport in CF epithelial cells by transfection of CFTR cDNA. Of particular interest will be the mechanism by which the biochemical activity of CFTR regulates, either directly or indirectly, the absorption and secretion of chloride ion. In terms of potential therapy for CF, the system we describe offers a means to screen for pharmacological compounds that interact with CFTR, and to test protein replacement and gene therapy as approaches in the treatment of CF. □

Received 27 July; accepted 31st August 1990.

1. Boat, T., Welsh, M. & Beaudet, A. in *The Metabolic Basis of Inherited Disease* (eds Scriver C., Beaudet A., Sly, W. & Valle, D.) 2649–2680 (McGraw Hill, New York, 1989).
2. Quinton, P. *Clin. Chem.* **35**, 726–730 (1989).
3. Li, M. *et al. Nature* **331**, 358–360 (1988).
4. Frizzell, R., Rechkemmer, G. & Shoemaker, R. *Science* **233**, 558–560 (1986).
5. Welsh, M. *Science* **232**, 1648–1650 (1986).
6. Rommens, J. *et al., Science* **245**, 1059–1065 (1989).
7. Riordan, J. *et al., Science* **245**, 1066–1073 (1989).
8. Kerem, B.-S. *et al. Science* **245**, 1073–1080 (1989).
9. Brosius, J. *Gene* **27**, 151–160 (1984).
10. Hawley, D. K. & McClure, W. R. *Nucleic Acids Res.* **11**, 2237–2255 (1983).
11. Cohen, S. N., Chang, A. C. Y., Boyer, H. W. & Heling, R. B. *Proc. natn. Acad. Sci. U.S.A.* **70**, 3240–3244 (1973).
13. Elroy-Stein, O., Fuerst, T. R. & Moss, B. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6126–6130 (1989).
14. Greenberger, L. M., Williams, S. S. & Horowitz, S. B. *J. biol. Chem.* **262**, 13685–13689 (1987).
15. Nash, B. & Tate, S. S. *J. biol. Chem.* **259**, 678–684 (1984).
16. Richard, N. D., Aldwin, L., Nitecki, D., Gottesman, M. & Pastan, I. *Biochemistry* **27**, 7607–7613 (1988).
17. Fuerst, T. R., Niles, E. G., Studier, W. & Moss, B. *Proc. natn. Acad. Sci. U.S.A.* **83**, 8122–8126 (1986).
18. Kohler, G. & Milstein, C. *Nature* **256**, 495–497 (1975).
19. Cheng, S. H. *et al. EMBO J.* **7**, 3845–3855 (1988).
20. Harlow, E., Crawford, L. V., Pim, D. C. & Williamson, N. M. *J. Virol.* **39**, 861–869 (1981).
21. Wurzner, R., Oppermann, M., Zierz, R., Baumgarten, H. & Gotze, O. *J. Immun. Meths.* **128**, 231–237 (1980).
22. Kawata, M. *et al. J. biol. Chem.* **264**, 15688–15695 (1989).
23. Muriakarmi, H. & Masui, H. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3464–3468 (1980).
24. Lieber, M., Mazetta, J., Nelson-Rees, W., Kaplan, M. & Todaro, G. *Int. J. Cancer* **15**, 741–747 (1975).
25. Rich, D. P. *et al. Nature* **347**, 358–363 (1990).
26. Sambrook, J., Fritsch, E. F. & Maniatis, T. in *Molecular Cloning, A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1989).
27. Stoker, N. G., Fairweather, N. F. & Spratt, B. G. *Gene* **18**, 335–341 (1982).
28. Laemmli, U. K. *Nature* **227**, 680–685 (1970).
29. Tarentino, A. L., Gomez, C. M. & Plummer, T. H. *Biochemistry* **24**, 4665–4671 (1985).
30. Feigner, P. L. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 7413–7417 (1987).
31. Harlow, E. & Lane, D. In *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1988).
32. Harvey, R., Hehir, K. M., Smith, A. E. & Cheng, S. H. *Molec. cell. Biol.* **9**, 3647–3656 (1989).
33. Mole, S. & Lane D. *J. Virol.* **54**, 703–710 (1985).

ACKNOWLEDGEMENTS. We thank G. White for oligonucleotides; G. Horn and D. Dearborn (Cleveland) for advice and reagents in the early stages of this work; B. Moss (NIH) for the plasmid pTM-1; H. de Jonge (Rotterdam) for advice, and for exchanging antibodies; C. O'Riordan, D. Souza, J. McPherson and R. Douglas (Genzyme) for advice and help, and B. Herrera for secretarial assistance. This work was supported in part by the CF Foundation and the National Heart, Lung and Blood Institute.

Printed in Great Britain by Turnergraphic Limited, Basingstoke, Hampshire

Exhibit 2

Maximizing Gene Expression

Edited by

William Reznikoff
University of Wisconsin
Madison, Wisconsin

Larry Gold
University of Colorado
Boulder, Colorado

Butterworths
Boston London Durban Singapore Sydney Toronto Wellington

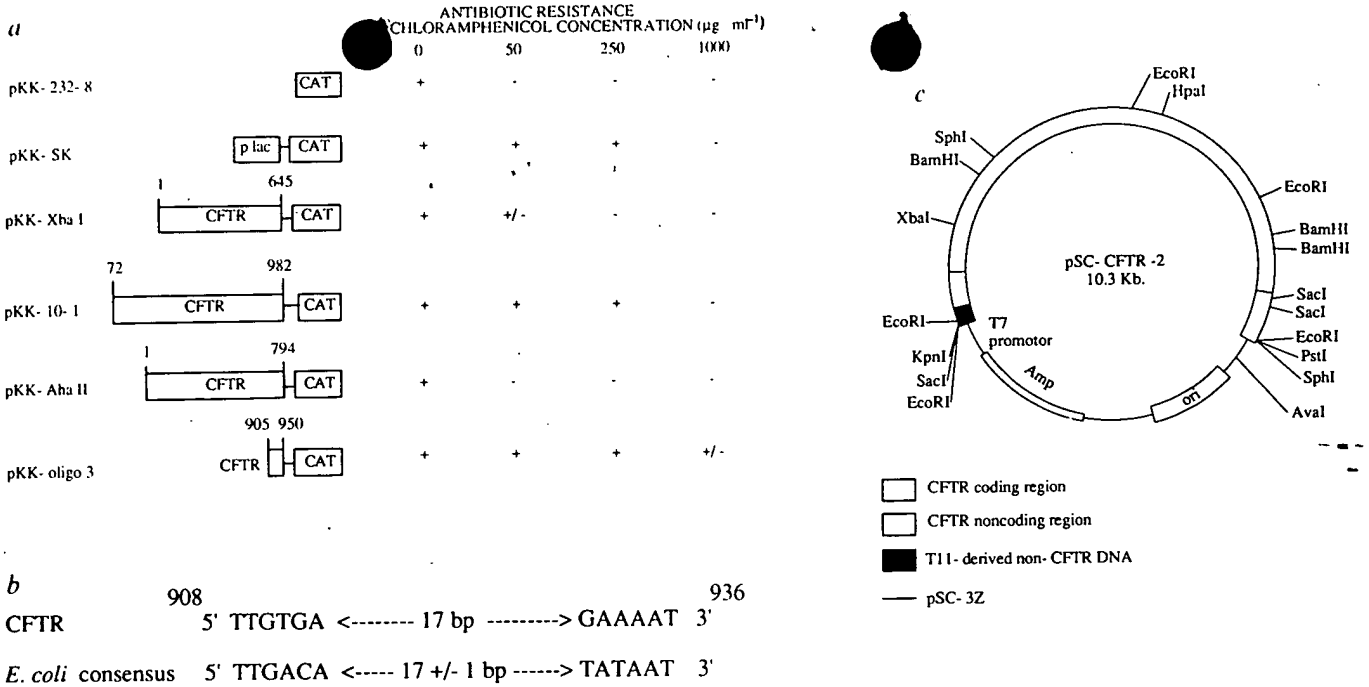


FIG. 1 Identification of a functional *E. coli* promoter within the CFTR cDNA and assembly of the complete CFTR open reading frame. *a*, Plasmid clones used to identify promoter sequences within the CFTR cDNA by activation of a promoterless CAT gene[10]. *E. coli* cells containing the constructs were challenged with increasing concentrations of chloramphenicol. Control plasmids contained either no insert (pKK232-8) or the β-galactosidase promoter (plac) from pBluescript SK⁻ (Stratagene). Nucleotide coordinates for the CFTR gene fragments are as given by Riordan *et al.*[7]. Symbols: +, growth to confluence; +/−, slow growth; −, no growth. *b*, Alignment of the promoter sequence within the CFTR cDNA with the consensus *E. coli* promoter sequence [11] *c*, Map of the low-copy-number plasmid pSC-CFTR-2, which contains the first 5.5 kilobases of the CFTR sequence of Riordan *et al.*[7], including the entire open reading frame.

METHODS. DNA manipulations were according to published procedures[26]. All CFTR plasmids were constructed using the partial CFTR cDNA clones T11, T16-1, T16-4.5 C1-1/5 and 10-1 (ref. 7), (obtained from the American Type Culture Collection), except for PKK-oligo3, which contains synthetic oligonucleotides corresponding to nucleotides 905–950 of the CFTR cDNA. The CFTR open reading frame was reconstructed in a plasmid (pSC-3Z) containing the pSC101 origin of replication from plasmid pLG338[11,27] combined with the β-lactamase gene and polylinker of pGEM-3Z (Promega). Base pairs (bp) 1–645 of pSC-CFTR-2 were derived from T16-1, bp 646–1,817 from T16-1 and bp 1,818–5,576 from T16-4.5. The CFTR open reading frame within pSC-CFTR-2 was completely sequenced by the dideoxy sequenase method. To construct a ΔF508 variant of pSC-CFTR-2 (pSC-CFTR-2ΔF508), the sequence between nucleotide positions 1,504 and 2,463 was replaced with the corresponding segment from CI-1/5, which contains the ΔF508 mutation.

standard procedures[18].

Figure 3*a* shows the immunoprecipitation of the cell-free product made in the presence of membranes and of the protein synthesized in HeLa cells transfected with CFTR cDNA. Bands A and B are efficiently immunoprecipitated by polyclonal antibody Ex13 (lanes 3 and 7). Neither preimmune serum nor immunoaffinity purified antibody raised against an irrelevant fusion protein (β-galactosidase-c-fyn; ref. 19) recognize the protein (lanes 1 and 2). Furthermore, the interaction is reversed by addition of excess β-galactosidase-exon 13 fusion protein (lanes 4, 5, 8), but not by an excess of irrelevant fusion protein (lane 6). These data show that the antibody specifically recognizes both the proteins synthesized *in vitro* and the recombinant cell-derived proteins.

Figure 3*b* shows immunoprecipitation of CFTR using two monoclonals (13-1 and 13-2) and immunoaffinity-purified polyclonal antibody Ex10. Band B is recognized efficiently and specifically. In all cases, addition of the appropriate fusion protein inhibits immunoprecipitation. An irrelevant control monoclonal antibody (raised against SV40 T-antigen; ref. 20) does not recognize the protein. Epitope mapping of the two monoclonals[21] indicates that they recognize non-overlapping regions within exon 13 (not shown). Because two polyclonal antibodies and two distinct monoclonals raised against predicted CFTR sequences recognize the protein made *in vitro* and *in vivo*, there can be little doubt that the 130/135 K protein represents the product of the CFTR cDNA.

Our initial attempts to immunoprecipitate CFTR protein from metabolically labelled nonrecombinant cells were unsuccessful, probably because the amount of protein in these cells is small. We attempted to overcome the problem by phosphorylating the immunoprecipitated protein using the catalytic subunit of protein kinase A and [γ^{32}P]ATP (ref. 22). Figure 4 shows proteins synthesized *in vitro* and labelled with ^{32}P after immunoprecipitation with the polyclonal antibody Ex13. Lane 1 shows that a protein migrating in the position of band B is labelled in immunoprecipitates of material synthesized *in vitro* in the presence of canine microsomal membranes and full-length CFTR cDNA. More rapidly migrating proteins are detected in immunoprecipitates of the cell-free product made in the absence of membranes (lane 2) or under the direction of the truncated cDNA (lane 3). These results indicate that CFTR synthesized *in vitro* is a good substrate for protein kinase A when it is present in immunoprecipitates.

Immunoprecipitates of extracts of nonrecombinant cells were labelled in the same way. Lane 5 shows that ^{32}P-labelled material comigrating with band B was detected in immunoprecipitates of T84 cells[23] treated with protein kinase A. A more highly phosphorylated, diffuse band of $M_r \sim 150$ K (band C) was also present. T84 cells are a human colon carcinoma cell line known to contain relatively large amounts of CFTR messenger RNA[7]. Bands comigrating with B and C were not detected in Panc-1 cells[24] (lane 4), mouse 3T3 cells (lane 6) or 27Sk human fibroblasts (not shown). It is believed that these cells do not express CFTR[7]. Bands comigrating with bands B and C were also detected in similar experiments using extracts of T84 cells that had been reacted with the antibodies Ex10, 13-1 and 13-2 (not shown).

The distribution in different cell types and the finding that four antibodies against CFTR sequences recognize material of similar mobility, suggests that the bands B and C labelled in T84 cell immunoprecipitates are CFTR. To confirm this, we

Reprinted from Nature, Vol. 347, No. 6291, pp. 382–386, 27th September, 1990

EXHIBIT I

Expression and characterization of the cystic fibrosis transmembrane conductance regulator

Richard J. Gregory, Seng H. Cheng, Devra P. Rich*, John Marshall, Sucharita Paul, Kathleen Hehir, Lynda Ostedgaard*, Katherine W. Klinger†, Michael J. Welsh* & Alan E. Smith‡

Genzyme Corporation and † IG laboratories Inc., One Mountain Road, Framingham, Massachusetts 01701, USA
* Howard Hughes Medical Institute, Departments of Internal Medicine and Physiology and Biophysics, University of Iowa, College of Medicine, Iowa City, Iowa, USA

CYSTIC fibrosis (CF) is a common lethal genetic disease that manifests itself in airway and other epithelial cells as defective chloride ion absorption and secretion[1,2], resulting at least in part from a defect in a cyclic AMP-regulated, outwardly-rectifying Cl^- channel in the apical surface[3–5]. The gene responsible for CF has been identified and predicted to encode a membrane protein termed the CF transmembrane conductance regulator (CFTR)[6–8]. Identification of a cryptic bacterial promoter within the CFTR coding sequence led us to construct a complementary DNA in a low-copy-number plasmid, thereby avoiding the deleterious effects of CFTR expression on *Escherischia coli*. We have used this cDNA to express CFTR *in vitro* and *in vivo*. Here we demonstrate that CFTR is a membrane-associated glycoprotein that can be phosporylated *in vitro* by cAMP-dependent protein kinase. Polyclonal and monoclonal antibodies directed against distinct domains of the protein immunoprecipitated recombinant CFTR as well as the endogenous CFTR in nonrecombinant T84 cells. Partial proteolysis fingerprinting showed that the recombinant and non-recombinant proteins are indistinguishable. These data, which establish several characteristics of the protein responsible for CF, will now enable CFTR function to be studied and will provide a basis for diagnosis and therapy.

Initial attempts to assemble a full-length CFTR coding sequence resulted in extensive rearrangements in the DNA clones obtained. This, together with the reported absence of full-length clones in the original isolates, made us consider whether transfected bacteria might be selecting against the full-length CFTR cDNA. Although the CFTR DNA sequence includes some repeated elements, they are not more frequent or extensive than is typical for other stable DNA sequences. This argued against intrinsic DNA instability as the cause of the problem but instead suggested that the presence of the full-length cDNA was toxic to cells. Such toxicity generally results from the inappropriate expression of cDNA and could be due to the presence of cryptic promoters within the DNA sequence. To search for any such transcription initiation sites, segments of CFTR were placed upstream of a promoterless gene encoding chloramphenicol acetyl transferase (CAT), transfected into *E. coli*, and the bacteria challenged with chloramphenicol[9]. By this means we identified an active promoter sequence beginning at residue 908 in exon 6B (Fig. 1*a*). This sequence shows good homology with the consensus *E. coli* promoter sequence, having 9 of 13 residues identical, including a highly conserved T residue at the 3′ end of the −10 box (Fig. 1*b*)[10].

To circumvent this problem of cDNA toxicity, we reduced the gene dosage of the cDNA by assembling a full-length CFTR open reading frame in a vector containing a low copy number of DNA replication originally derived from the plasmid pSC101 (ref. 11) (pSC-CFTR-2, Fig. 1*c*). A second plasmid pSC-CFTR-2ΔF508 was constructed which contains a deletion corresponding to amino-acid residue 508, the site of a common CF mutation[7]. The resulting clones had no deleterious effect on bacterial growth and were stable. DNA sequence analysis of the complete coding sequence revealed a sequence identical to that reported by Riordan *et al.*[7], with the exception that the base at position 1,990 is C rather than A.

To examine the protein product of the CFTR cDNA, we constructed low-copy-number vectors (pTM-CFTR-3 and pTM-CFTR-3ΔF508) with the protein coding sequence immediately downstream of a bacteriophage T7 promoter and an encephalomyocarditis (EMC) virus internal ribosome entry sequence and followed by a T7 transcriptional terminator[12,13]. The T7 and EMC sequences were isolated from the plasmid PTM-1 (T. Mizukami, O. Elroy-Stein and B. Moss, personal communication). After transcription *in vitro* using T7 RNA polymerase, the resulting RNA was translated in a reticulocyte lysate in the presence of [^{35}S]methionine. Figure 2*a* shows an autoradiograph of the cell-free product after gel electrophoresis. A major band migrating with a relative molecular mass of 130,000(M_r 130K) was detected (band A, lane 2). When a truncated transcript generated by linearization at the *Tth*111I site (amino acid 1,336) was translated, the M_r of the cell-free product (lane 5) was, as expected, reduced by ~10 K. The M_r of the *in vitro* product is somewhat smaller than that expected for a protein consisting of 1,480 amino acids. But Greenberger *et al.*[14] have reported that the related multiple drug resistance protein also migrates faster than expected on polyacrylamide gels.

In vitro translation was enhanced by the presence of detergent (0.5% Triton X-100)[15] in the reticulocyte cell-free extract (lane 6). Furthermore, boiling the protein in sample buffer before electrophoresis dramatically reduced the amount of material migrating at 130 K and increased the material at the top of the gel (lane 8). Both these observations are reminiscent of results[14–16] with other integral membrane proteins and are probably a consequence of the presence of several hydrophobic transmembrane sequences within the polypeptide.

When translation in the reticulocyte lysate was performed in the presence of canine microsomal membranes, the mobility of the cell-free product decreased to correspond to M_r ~ 135 K (band B, lane 3). This implies that band A is modified by the membranes by addition of core carbohydrate. Consistent with this, when band B was treated with detergent and *N*-Glycanase its mobility increased to comigrate with band A (lane 4). Translation of the ΔF508 mutant cDNA gives a primary *in vitro* product indistinguishable from wild type (lane 7).

To examine the expression of CFTR in transfected cells we used the vaccinia virus/T7 RNA polymerase hybrid expression system[13,17]. HeLa cells were infected with a vaccinia virus encoding the RNA polymerase of bacteriophage T7 and subsequently transfected with the vector pTM-CFTR-3 or control plasmids. Figure 2*b* shows a polyacrylamide gel of the proteins labelled with [^{35}S]methionine after synthesis in infected and uninfected cells. Material co-migrating with bands A and B is present in infected cells transfected with the pTM-CFTR-3 vector, but not in those transfected with the control plasmid pTM-1 or in uninfected cells. Treatment with *N*-Glycanase resulted in a mobility shift of the upper of the two bands to generate material comigrating with band A (data not shown). Although not visible in Fig. 3, another band of M_r 150 K is also present in small amounts in the cells transfected with CFTR cDNA (see later).

To characterize further the CFTR gene product we raised antisera in rabbits against fusion proteins containing the amino-acid sequences encoded by exon 13 (polyclonal antibody Ex13) or exon 10 (polyclonal antibody Ex10) fused to β-galactosidase. Exon 13 encodes the unique, polar, R domain[7]. Exon 10 encodes a portion of the first nucleotide-binding fold, including the site of the common CF-related deletion, ΔF508[7]. Mouse monoclonal antibodies to the exon 13 fusion protein (monoclonal antibodies 13-1 and 13-2) were also produced, screened and cloned using

‡ To whom correspondence should be addressed.

en the new data suggesting that
ntain strong protein homologies?
lation in eukaryotes be as forbid-
hetic times for long transcripts
e nuclear membrane allow cou-
gg McMuffin that a future book
to simplify many of our views.
ple, especially the contributors.
outors during the review period
from experts. We also enjoyed

CHAPTER
1

E. Coli Promoters

William S. Reznikoff
William R. McClure

The 2,000 plus genes in the *Escherichia coli* chromosome are frequently organized into groups or operons that are expressed as independent units of transcription. For each of these units there is a DNA sequence, termed a *promoter*, that signals the start of the transcript (i.e., it stimulates the RNA polymerase–DNA interaction that leads to transcription initiation) and a DNA sequence that signals the stop of the transcript (stimulating transcripting termination). This chapter analyzes the process of transcription initiation and the DNA signals that stimulate this event.

To a first approximation, promoters are DNA sequences that are recognized by RNA polymerase holoenzyme such that it catalyzes the transcription initiation process. It is believed that the rate of transcription initiation is, in the simplest cases, dictated by the nature of the DNA sequence composing the promoter. This chapter first describes how promoters can be defined and then discusses the mechanism of transcription initiation and how promoter structure relates to this mechanism. However, we shall also discuss several important complications of this definition. An appreciation of these complications

We thank many colleagues and co-workers for communicating results prior to publication. Our research on transcription initiation is supported by the National Institutes of Health (NIH) (GM 19670 to WSR and GM 30375 to WRM) and the 3M Foundation (to WSR).

is important for a critical understanding of the process of transcription initiation. For example:

1. The frequency of transcription initiation at many promoters is regulated by the interaction of other proteins with defined target sequences. In some cases the regulation is negative (as in repressor-operator interactions), while in others it is positive (in which case the bound regulatory protein enhances the transcription initiation rate inherent in the interaction of RNA polymerase with the particular promoter). Other types of regulatory controls also exist.
2. Some DNA sequences that appear to act as promoters *in vitro* apparently do not act so *in vivo*, thus complicating attempts to define *in vitro* promoters. The name *tight binding sites* has been used to describe some of these sites. In at least one case such a sequence overlaps a functional promoter, and this sequence may have an important regulatory role.

1.1 DEFINING PROMOTERS

Promoters have been defined by diverse types of evidence. This diversity has been due in part to the different approaches that have been used, which in turn has been due to (1) the properties of different systems that make them more or less amenable to different tactics, (2) the technical predilections of the laboratories studying given systems, and (3) historical accidents. The conclusion we draw from the following analysis of some of these approaches is that a final definition of a promoter requires a combination of *in vivo* and *in vitro* experiments.

1.1.1 Genetic Analysis

A primary method for defining promoters has been the isolation and characterization of mutations that alter promoter function. A classical approach to this method was pursued by Beckwith and his colleagues in the analysis of the *lac* promoter (Scaife and Beckwith 1967; Silverstone et al. 1970; Beckwith et al. 1972; Arditti et al. 1973; Hopkins 1974; Beckwith 1981). Mutations were isolated that altered (either decreased or increased) the levels of expression for all three *lac* genes. The detailed properties of these mutations also provided criteria for defining promoter mutations. The levels of expression were coordinately altered. The mutations were *cis*-dominant. The mutations were not suppressed by either nonsense suppressors or the polarity suppressor SuA (a ρ mutation). The mutations mapped at the start of the operon. Figure 1–1 shows the nucleotide sequence changes found for various *lac* promoter mutations. One can see that they are clustered within a 50 bp sequence. As is described later, these mutations define characteristics of the promoter sequence that also fit with other types of analyses such as the compiling and

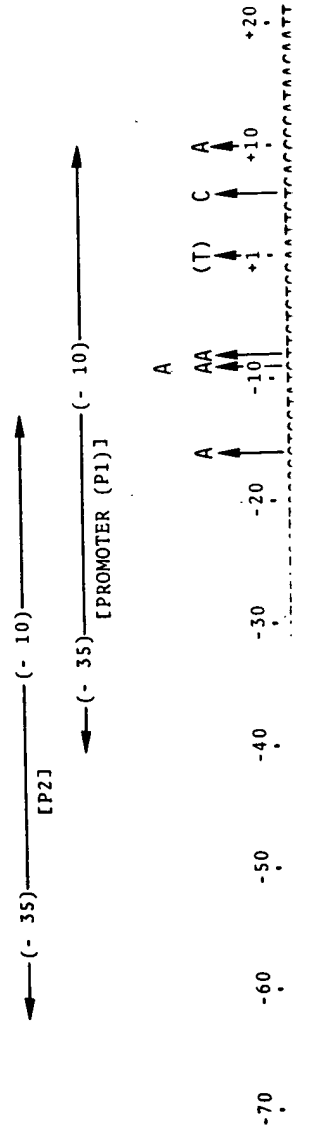


FIGURE 1-1 The *lac* promoter and operator. One strand of the *lac* promoter and operator DNA sequence is presented with the bases numbered as follows: +1 = mRNA startsite; positive numbers = downstream sequences; negative numbers = upstream sequences. Mutations that affect *lac* promoter activity include those that enhance *lacP* expression (indicated above the line) and decrease *lacP* expression (below the line). Also shown is a change at +1 |A→ (T)| that, although it enhances *lac* expression, does not do so by altering the activity of the *lacP* sequence. These mutations have been described before by Reznikoff (1984) and Kunkle (1984). Shown above the sequence are the approximate boundaries of the *lac* promoter and its – 35 and – 10 regions and the approximate location of the overlapping RNA polymerase binding site P2 (Peterson and Reznikoff 1984a). It should be noted that most of the promoter mutations affect the similarity of the sequence in the – 10 and – 35 *lacP* regions to the canonical sequences or affect the spacing of the regions in a manner consistent with their predicted importance (see Figure 1–3). Moreover, many of the mutations in the *lacP* – 35 region also happen to be in the – 10 region of P2. Also shown in the figure are the sequence changes that have been found in the *lac* operator associated with various *lacO*^{c} mutations (see Reznikoff 1984 for a summary of this information).

comparing of all known promoter sequences (see Figure 1–3) and the results of chemical and enzymatic probe experiments.

The *lac* system was particularly amenable to this type of analysis because it was technically easy to select and/or screen for mutations, which resulted either in an increase or a decrease in *lac* expression. This advantage can be generalized to any *E. coli* transcription unit and even to systems in other organisms, thanks to the development of procedures that allow the generation of fusions of the system of choice to the *lacZ* gene.

Although the *lac* studies were very profitable in facilitating the genetic analysis of promoter structure, they also exemplified one of the most important complications of this type of analysis. Mutations that alter the DNA target site for a positive regulatory protein (in this case, the CAP-cAMP complex) resemble promoter (RNA polymerase target site) mutations in all of the listed criteria (Beckwith et al. 1972; Hopkins 1974). One method for distinguishing mutations that decrease RNA polymerase recognition of the promoter from mutations that decrease positive activator-DNA interaction has been outlined by Beckwith (1981). It involves an examination of the mutant's residual *lac* expression level for its sensitivity to the presence or absence of the positive activator. Another approach involves the use of specific *in vitro* assays for alterations in RNA-polymerase-DNA and positive activator-DNA interaction. Finally, the location of the sequenced mutation often provides some insight into the step involved. However, all these approaches make simplifying assumptions about the organization of these recognition sites (i.e., are they distinct or do they overlap?) and the mechanism of positive regulator action (see Section 1.5.2).

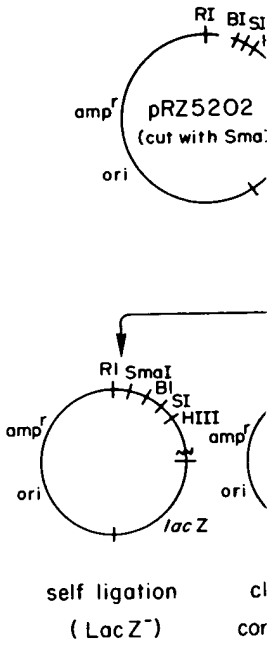


FIGURE 1–2 A typical p
structed by Munson (1983)
W200 (Yu et al. 1984) such
fragment is cloned in the c
*Bam*HI (BI), *Sal*I (SI), or *H*

1.1.2 Promoter Cloning Vehicles

The advent of recombinant DNA technology, coupled with the development of operon fusion techniques, has led to the construction of promoter cloning vehicles. These vehicles have provided new *in vivo* approaches for the identification of DNA sequences containing promoters. They are designed to have unique cloning sites located upstream from a gene encoding an easily assayable and/or selectable function, such as *lacZ* (β-galactosidase), *galK* (galactokinase), *cat* (chloramphenicol acetyltransferase), or *tetA* (tetracycline resistance) (An and Friesen 1979; West et al. 1979; Casadaban and Cohen 1980; McKenney et al. 1982; Mandecki and Reznikoff 1982; Bertrand et al. 1984; Wertman et al. 1984; Munson et al. 1984). In the parent vector this gene has a null expression phenotype since no promoter exists to program its messenger synthesis.

An example of such a vector and its use is shown in Figure 1–2. The procedure involves the cloning of previously mapped fragments, or shotgun cloning of fragments, to be subsequently mapped into the vehicle's cloning site(s). Fragments that contain a promoter and that are situated in the correct orientation turn on the expression of the indicator gene. These vehicles are useful for identifying which of a set of restriction fragments contains a promoter, determining the orientation of a promoter within a given fragment,

and obtaining a qualita
Quantitative measureme
variables, such as copy
differences due to the di
initiation frequencies du
ters 4, 5, and 7).

Promoter cloning v
detailed genetic analyse
They are directly amen
the other boundary of a
side of the promoter
Reznikoff 1984). They
point mutations. These
genic treatments or by n
sis. For instance, LeCler
(1984) have shown that
be used to screen for mu
ing plaque color on inc

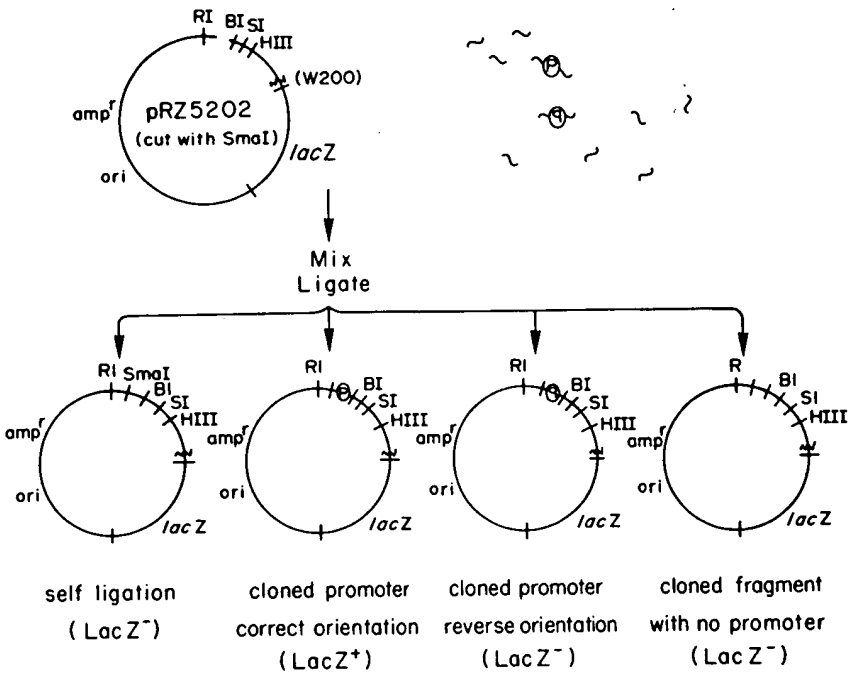


FIGURE 1–2 A typical promoter cloning vehicle. The plasmid pRZ5202 was constructed by Munson (1983) from pBR322 into which was cloned the *trp-lac* fusion W200 (Yu et al. 1984) such that *lacZ* is not expressed unless a promoter-containing fragment is cloned in the correct (clockwise) orientation into the *Eco*RI (RI), *Sma*I, *Bam*HI (BI), *Sal*I (SI), or *Hind*III (HIII) sites.

and obtaining a qualitative estimate of a given promoter's level of activity. Quantitative measurements of activity require the control of several additional variables, such as copy number variations of the vector, mRNA stability differences due to the different mRNA 5′ ends, and differences in translation initiation frequencies due to varying mRNA secondary structures (see Chapters 4, 5, and 7).

Promoter cloning vehicles are particularly useful in their facilitation of detailed genetic analyses of promoters (and other genetic regulatory signals). They are directly amenable to the generation of deletions that define one or the other boundary of a promoter. One merely uses a restriction site on one side of the promoter to generate *Bal*31 or S1 nested deletions (Yu and Reznikoff 1984). They also facilitate the isolation and sequence analysis of point mutations. These mutations can be generated either by random mutagenic treatments or by means of synthetic oligonucleotide-directed mutagenesis. For instance, LeClerk and Istock (1982), Kunkel (1984), and Munson et al. (1984) have shown that the phage M13 variants mp2, mp8, and mRZ361 can be used to screen for mutations in the *lac* control elements merely by examining plaque color on indicator plates; interesting candidates can be directly

sequenced using the known dideoxy-mp8 methodologies (some of the mutations shown in Figure 1–1 were isolated by this approach). Other mp8-like constructs have been developed that allow similar analyses of other controlling elements (Wertman et al. 1984).

1.1.3 Determining the 5′ End of the mRNA

The determination of the precise 5′ end of the mRNA is a critical piece of information that allows the detailed structural analysis of promoters. It permits the alignment of the sequence information for different promoters, as well as the comparison of mutational changes and the results of chemical and enzymatic probe experiments. The typical approach for the *in vitro* determination of the 5′ mRNA end is to program a transcription reaction from a defined DNA fragment known to contain the promoter of interest and then to analyze the resulting products either by their size and orientation or in terms of their actual sequence. In some cases [such as *lac* (Reznikoff et al. 1982; McClure et al. 1982)] multiple transcripts are programmed from the fragment, and the experimenter emphasizes the major transcript from among those that have the correct orientation. This approach raises questions such as: Has the correct transcript been identified? Would changing the topology of the template to resemble more closely the *in vivo* situation change the types and amounts of products made? Does the presence of other transcripts signify other RNA polymerase binding sites, and if so, do they have a physiological role? Moreover, in a few cases (e.g., *lacZ* and *lacI*) in which the wild-type promoter is weak, mutant and not wild-type templates were initially used for determining the startsite (Maizels 1973; Steege 1977). Since, as we shall point out, mutations can generate new transcription startsites, these experiments could have misidentified the wild-type 5′ end.

These kinds of observations suggest that it is critical to determine the *in vivo* 5′ end of the message. The *trp* operon is an example of a simple single start point system in which the *in vivo* and *in vitro* determinations are in agreement (Squires et al. 1976). More complex systems in which *in vitro* and *in vivo* results are the same include *gal* (two differently controlled starts in the same orientation) and the Tn10 *tet* region (multiple and divergent starts (Aiba et al. 1981; Bertrand et al. 1983; Hillen and Schollmeir 1983). Other examples are also known [see Hawley and McClure (1983a) for some examples].

The *lac* system is not as straightforward. *In vivo*, the CAP-cAMP stimulated wild-type promoter programs a transcript that is the same as that found *in vitro* (Munson 1983; Peterson and Reznikoff 1984a; Cannistraro and Kennell 1985). However, the wild-type promoter also programs other transcripts *in vitro* that are not evident *in vivo* (Reznikoff et al. 1982; McClure et al. 1982; Peterson and Reznikoff 1984a). Finally, the transcript programmed by one mutant, which elevates *lac* expression, P′115, starts approximately 13 bp downstream from the normal startsite (Maquat and Reznikoff 1980; Peterson and Reznikoff 1984b). Thus, it was fortuitous that the original *in vitro* tran-

scription studies (Maizel
activates transcription fr
nate site.

In spite of the cauti
promoter, in most syste
vitro with the *in vivo* star
aspects of promoter func
fidelity in the *in vitro* re
catalytic and the regulat
may have a significant ap
control of transcription i
in this chapter.

1.1.4 Binding Experim

Transcription initiation
binding of RNA polyme
a stable open complex,
detail in Section 1.3). A
ments, have been used t
dures usually rely on the
open complex with RN
physiological temperatu
promoters, but normall
(<0.3 M Na^+). The con
inactivation of RNA pc
form such complexes h
binding experiments, al

The formation of R
as evidence that a partic
rate of forming this c
Although both seem to
plification that is probal
tested. Caution is also ir
not result in high filter-l

Two classes of seque
that, however, do not a
tight binding sites (Kad
sequences lie at appare
promoter locations (as j
complexes formed at sc
RNA synthesis. This ty
than bona fide promote
they were promoters an
nates against these spu

scription studies (Maizels 1973; Majors 1975) utilized a mutant template that activates transcription from the wild-type startsite rather than from an alternate site.

In spite of the cautions raised by experiences gained in studying the *lac* promoter, in most systems studied there is a general concurrence of the *in vitro* with the *in vivo* startsite determinations. This finding suggests that many aspects of promoter function and promoter recognition occur with rather high fidelity in the *in vitro* reactions and that cautious extrapolation of both the catalytic and the regulatory properties of RNA polymerase deduced *in vitro* may have a significant application to achieving an understanding of the *in vivo* control of transcription initiation. Such an analysis is presented subsequently in this chapter.

1.1.4 Binding Experiments

Transcription initiation occurs through a series of steps that include the binding of RNA polymerase, the isomerization of the bound complex to form a stable open complex, and the initiation of transcription (described in more detail in Section 1.3). A variety of protocols, which are called *binding experiments*, have been used to define promoterlike elements in DNA. These procedures usually rely on the ability of a particular DNA fragment to form a stable open complex with RNA polymerase. The complexes form only at close to physiological temperatures (the minimum temperature differs for different promoters, but normally exceeds 15°C) and moderate salt concentrations (<0.3 M Na^+). The complexes are relatively stable and are resistant to the inactivation of RNA polymerase by polyanions like heparin. The ability to form such complexes has been measured primarily by nitrocellulose filter-binding experiments, although electron microscopy has also been used.

The formation of RNA-polymerase-DNA open complexes is usually taken as evidence that a particular DNA sequence contains a promoter and that the rate of forming this complex is related to the strength of the promoter. Although both seem to be reasonable assumptions, the former is an oversimplification that is probably not true in many cases, and the latter is still being tested. Caution is also indicated by the fact that some bona fide promoters do not result in high filter-binding efficiency in these assays.

Two classes of sequences bind to RNA polymerase to form open complexes that, however, do not appear to be promoters. The first class has been called tight binding sites (Kadesch et al. 1980; Melançon et al. 1982). These binding sequences lie at apparently random locations. They are unrelated to known promoter locations (as judged by genetic and physiological criteria). The open complexes formed at some of these sites are surprisingly capable of initiating RNA synthesis. This type of observation suggests that many sequences other than bona fide promoters may be recognized by RNA polymerase *in vitro* as if they were promoters and raises the question as to what, if anything, discriminates against these spurious sites *in vivo*.

The second class is best represented by the P2 sequence in the *lac* controlling elements (see Figures 1–1 and 1–7) (Peterson and Reznikoff 1984b; Malan and McClure 1984; Spassky et al. 1984). RNA polymerase will form open complexes in the absence of CAP-cAMP, with DNA containing the *lac* controlling elements, but the relevant sequence is not the *lac* promoter. Rather, it is a sequence that is displaced 22 bp upstream from, and thus overlaps with, the *lac* promoter. This is, in fact, the sequence that programs one of the additional transcripts *in vitro*, although there is no evidence that this transcript is synthesized *in vivo* (if it is synthesized it must be so unstable that it is not detectable by S1 mapping experiments and cannot program gene expression when incorporated into promoter cloning vehicles). To gain a perspective as to how this type of site can confound an analysis of a promoter, one merely needs to re-analyze the RNA-polymerase-*lacP* binding experiments reported by Reznikoff (1976). It is now clear that the reported perturbation of RNA-polymerase-*lacP* binding by *lacP* −35 region mutations was due to the fact that these modifications are in the −10 region of P2 and not due to their position in *lacP* (Peterson and Reznikoff 1984b).

The existence and properties of this binding site raise important questions. First, the failure to detect P2 programmed transcription *in vivo* suggests that P2 lacks a sequence determinant that controls transcription initiation *in vivo*. It is critical to determine the basis for the difference between *in vivo* and *in vitro* transcription initiation reactions and to determine what the missing sequence signal is in P2. Second, the location of the P2 binding site (overlapping both the CAP site and the promoter) suggests that it may play a role in the activation of transcription initiation by the CAP-cAMP complex (see Section 1.5.2) (McClure et al. 1982; Reznikoff et al. 1982; Malan and McClure 1984; Malan et al. 1984; Peterson and Reznikoff 1984b).

1.2 STRUCTURE ANALYSIS

The identification of a promoter site within a segment of DNA is the beginning of a process of successive refinements in elucidating its function both *in vivo* and *in vitro*. In this section we discuss those structural features of the promoter that are currently believed to be important to promoter function. First, the location of promoter mutations proves that DNA sequence is directly related to the diversity of function. Second, the patterns of DNA protection that result from RNA polymerase binding at the promoter suggest a unity in this interaction. Finally, the alterations of DNA conformation and supercoiling can in some cases have a large and specific effect on promoter utilization.

1.2.1 Comparative Sequence Analysis

The DNA sequences of 112 *E. coli* promoters were compiled by Hawley and McClure (1983a). A depiction of the consensus sequence for *E. coli* promoters is shown in Figure 1–3. [illegible] the promoters are show[n] that the consensus seq[uence] moter sequences added Siebenlist et al. 1980). S[ee] compilation can be stat[ed] tion of *E. coli* promoter[s] TA---T in the −10 regio[n] promoter. Other bases [near] either side or within the of this compilation is tha[t] by the location and id[entity] sequences of promoter n[ucleotide] homology to the consen[sus] base pair alterations th[at] up-promoter mutations. nonconsensus-to-noncon[sensus] ity of a hierarchy of re[gions] other rule that can be a[pplied] each known promoter s[equence] the three most highly co[nserved] −7 has frequently been [the] case. Indeed, the A res[idue] −7 T. There is conside[red] and the less stringent r[ule] has a match of at least o[ne] in that region.

In addition to the h[examer] the promoter has beer[n] spacing between the −3 [and] is ordinarily 17± 1 base

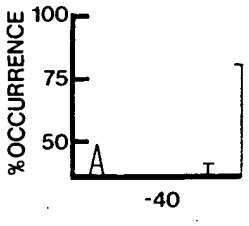


FIGURE 1–3 The conse[nsus] strand is displayed in hei[ght] promoters compiled by H[awley] level two standard deviati[ons] occurrence of bases at eac[h]

is shown in Figure 1–3. The most highly conserved bases in the two regions of the promoters are shown by the size of the individual letters. It is remarkable that the consensus sequence has changed only slightly with additional promoter sequences added to previous compilations (Rosenberg and Court 1979; Siebenlist et al. 1980). Several important conclusions derived from the current compilation can be stated in the form of rules for RNA polymerase recognition of *E. coli* promoters. It is clear that the TTG in the –35 region and the TA---T in the –10 region are the most highly conserved base pairs within the promoter. Other bases of varying significance in a statistical sense occur on either side or within these sequences. One rule that was deduced on the basis of this compilation is that the consensus sequence is best. The rule is supported by the location and identity of more than 100 promoter mutations. The sequences of promoter mutations show that base pair alterations that decrease homology to the consensus sequence are down-promoter mutations and that base pair alterations that increase homology to the consensus sequence are up-promoter mutations. The only exceptions to the rule are a handful of nonconsensus-to-nonconsensus base pair alterations that suggest the possibility of a hierarchy of recognition at each position within the promoter. Another rule that can be applied to all the sequenced *E. coli* promoters is that each known promoter shows a match of at least two out of three base pairs at the three most highly conserved sites within the –10 region. The T residue at –7 has frequently been referred to as "invariant." That seems not to be the case. Indeed, the A residue at –11 appears to be nearly as invariant as the –7 T. There is considerably more diversity in sequence in the –35 region, and the less stringent rule that applies is that every known *E. coli* promoter has a match of at least one out of three to the highly conserved TTG sequence in that region.

In addition to the homologies found in –35 and –10, another feature of the promoter has been shown to be important in function—namely, the spacing between the –35 and –10. The distance as conventionally expressed is ordinarily 17± 1 base pairs. Again, the evidence from mutations is that the

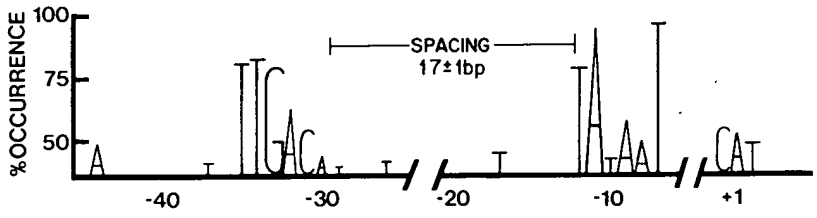


FIGURE 1–3 The consensus sequence for *E. coli* promoters. Each base on the top strand is displayed in height corresponding to its percent occurrence in the 112 promoters compiled by Hawley and McClure (1983). The baseline corresponds to a level two standard deviations (Poisson statistics) above that expected for random occurrence of bases at each position.

most frequently occurring spacing is the best. These studies have been performed *in vivo* and *in vitro*. As shown in Figure 1–3, there is another region of homology between *E. coli* promoters near the startsite of transcription. The three base pair weak homology surrounding the startsite of transcription has not been shown to be of any functional significance; that is, no mutations have been found in this region of any promoter that affect initiation frequency. Similarly, the distance between the −10 and the startsite can vary with about the same flexibility in number of base pairs as that observed between the −35 and −10. One cannot argue on the basis of no genetic evidence that there is no functional significance. However, at this time we will not emphasize the homology in this region for the discussion of promoter function and structure that follows.

We cannot deduce whether the −35 region is more or less important for function than the −10 region, despite previous claims (Okamoto et al. 1977) that some promoters can function without a −35 region. It has been noted in the past that many promoters, which are subject to positive control *in vivo*, have relatively poor homology to the consensus sequence in the −35 region. Raibaud and Schwartz (1984) have argued that there may be very little homology indeed to the −35 region consensus sequence for the class of positively controlled promoters. This type of analysis forces one to consider the role of the rather poor homology to the consensus sequence and how that affects promoter function. We can imagine two reasons why functional promoters might have relatively poor homology to the consensus. First, the sequence of the promoter frequently overlaps with other control sites such as operators and activator binding sites. Second, as we argue later, the rather poor homology to the consensus sequence is predicted to result in a weak promoter *in vivo*. The addition of a positive control element can convert the weak promoter to a strong promoter, thereby providing some essential adaptability to changes in cellular conditions.

We note briefly that the *E. coli* consensus promoter sequence is rather similar to the same sequences found in *Salmonella*, a close relative of *E. coli*, and even to *B. subtilis* promoters for the vegetative holoenzyme of that organism (Moran et al. 1982). A significant difference between the *B. subtilis* and *E. coli* promoters is that promoters from the gram positive organisms frequently are used at relatively high rates in *E. coli*, whereas the converse is not found (Moran et al. 1982).

1.2.2 Protection and Sequence

The limits of protection afforded the promoter DNA segment by the binding of the *E. coli* holoenzyme cover a region of ~65 base pairs. The pattern of protection from DNase I digestion is approximately the same for many promoters that have been investigated, although there may be some difference between strong promoters that form stable open complexes and weak promoters that form relatively less stable complexes. The chemical protection and

modification experiments
There are still not enough
unambiguous conclusion
moters with the same or
comparison was performe
technique in which the TA
temperatures using dimet
stranded cytosine residues
chemical modification wa
the melted region in the
cleavage of the end-labele
of this melted site. When
tures, a sharp transition
UV5 promoter and at ~1

A generalization from
on a limited number of pr
side of the helix, althoug
fashion that would wrap a
It is clear that an impor
electrostatic (Shaner et
bonding to base pairs is i
note that the priorities i
hierarchy both in positio
only three of the six bas
served, one could imagin
mutational evidence and
have shown that there
positions.

1.2.3 DNA Conformatic

For as long as promoters
tions of promoter functic
influencing gene expressi
tions is shown in Figure 1
that a change in structur
of events at a distant site
first, how far can the per
at a particular site, and
and the results at the dis
that the DNA context in
the consensus sequence
trivial example is that D
to allow polymerase bin
are not conserved. Th

modification experiments have been reviewed by von Hippel et al. (1984). There are still not enough examples of this type of experiment to allow the unambiguous conclusion that RNA polymerase recognizes the various promoters with the same or with different base pair recognition patterns. A comparison was performed using an alteration of the chemical modification technique in which the TAC and *lac*UV5 promoters were examined at various temperatures using dimethyl sulfate and exploiting its reactivity to single-stranded cytosine residues in DNA (Kirkegaard et al. 1983). In this case, the chemical modification was actually the result of exposing base pairs within the melted region in the promoter to the chemical reagent. Subsequent cleavage of the end-labeled promoter-containing DNA revealed the location of this melted site. When the experiment was performed at different temperatures, a sharp transition was observed that was centered at ~22°C for the UV5 promoter and at ~18°C for the TAC promoter.

A generalization from the protection experiments that have been performed on a limited number of promoters is that recognition occurs largely from one side of the helix, although the binding of RNA polymerase to the DNA in a fashion that would wrap around has not been excluded (Siebenlist et al. 1980). It is clear that an important favorable component in the overall binding is electrostatic (Shaner et al. 1983) and equally likely that specific hydrogen bonding to base pairs is involved (von Hippel et al. 1984). In that regard we note that the priorities in the consensus sequence noted earlier suggest a hierarchy both in position and for individual base pairs. For example, since only three of the six base pairs within each hexamer are very highly conserved, one could imagine that they are very important. At each position, the mutational evidence and some *in vitro* characterization (McClure et al. 1983) have shown that there are hierarchies of sequence preference at several positions.

1.2.3 DNA Conformation

For as long as promoters have been studyied *in vitro* there have been explanations of promoter function involving action at a distance or long-range forces influencing gene expression. A composite model that covers all these suggestions is shown in Figure 1–4 in the form of a cartoon. The central idea here is that a change in structure or energetic level at one site can affect the outcome of events at a distant site. At issue in all these models (real and whimsical) is first, how far can the perturbation in DNA be propagated once it has occurred at a particular site, and second, what is the mechanism of that propagation and the results at the distant site? We can state, without fear of contradiction, that the DNA context in and around the promoter region that extends beyond the consensus sequences certainly has an effect on promoter function. A trivial example is that DNA is required both downstream and upstream simply to allow polymerase binding, even though specific sequences in these regions are not conserved. That is not surprising since promoters are ordinarily

FIGURE 1–4 The cartoon shows the effect of action-at-distance on a scale corresponding to everyday experience. More realistic models corresponding to the molecular scale were recently discussed (Crothers and Fried 1983).

embedded within a continuous DNA helix inside the cell. The more important point at issue here for proponents of the action-at-a-distance explanations is what kinds of changes can occur in DNA and can they actually be shown to affect function directly? The fact that the spacer between the −35 and −10 can affect function both *in vivo* and *in vitro* suggests that twisting of the DNA could be important. Twisting may relate, in turn, to the supercoiled state of the DNA, and indeed as outlined below, supercoiling has several effects on promoter function. Originally proposed by Stefano and Gralla (1982 a, b) on the basis of spacer mutations within the $lacP^S$ promoter, the hypothesis that the spacer distance could be affected by twisting is a very appealing idea. In the same category, it has been shown from several lines of evidence that certain sequences of DNA can exhibit bends within the linear duplex (Kolb et al. 1983; Wu and Crothers 1984) or kinks such as the *Eco*RI binding site when complexed to that endonuclease (Frederick et al. 1984). Thus, the DNA molecule is, in fact, polymorphic; it can assume several different conformations in addition to the canonical B form.

Before getting carried away with possibilities raised by the diverse forms and structures that DNA can assume, a critical view of the matter dictates that we take into account the extent of the structural alterations and the rate and energies with which they occur. For most of the known perturbations, small changes such as twisting and bending are probably not significant either on a structural basis (± 5°) or a time scale (nsec) that relates to promoter function. Ultimately, one would like to find direct *in vitro* and genetic evidence that a change in DNA structure results in a change in function.

The only bona fide example of action-at-a-distance that has been demonstrated is the effect of supercoiling on promoter activity. We consider supercoiling an example of action-at-a-distance because DNA gyrase, the

enzyme in *E. coli* respo
on the DNA double heli
negative supercoils into
ous structure of the DN
of ATP (adenosine triph
a promoter site where it
DNA and initiate transc
First, transmission of
structure of the DNA t
convincing biochemica
in vivo and *in vitro* (Gel
distance fail on one or

1.3 STEPS IN TRANS

The fundamental moti
rate steps is the observa
that range from about
E. coli (i.e., rRNA ger
perhaps even less frequ
characterized by a lowe
once every few second
therefore, promoter fu
The remarkable featu
approximation, only o
frequencies. Thus, we
E. coli promoter sequ
directly to DNA sequer
the consensus sequenc
sequence a decrease i
occurring in living cell
mal function. For the e
certain promoters init
frequency.

1.3.1 Pathways to Op

The dissection of RNA
Zillig and co-workers
principal steps in initi
conversion of an inacti
a chain quickly. This sc
of Chamberlin (1974)
working model for RN

ıt-distance on a scale correspond-
corresponding to the molecular
3).

the cell. The more important
ı-at-a-distance explanations is
:an they actually be shown to
:r between the −35 and −10
gests that twisting of the DNA
n, to the supercoiled state of
coiling has several effects on
ıno and Gralla (1982 a, b) on
promoter, the hypothesis that
ıg is a very appealing idea. In
everal lines of evidence that
hin the linear duplex (Kolb et
; the *Eco*RI binding site when
t al. 1984). Thus, the DNA
e several different conforma-

:s raised by the diverse forms
iew of the matter dictates that
l alterations and the rate and
: known perturbations, small
bly not significant either on a
relates to promoter function.
) and genetic evidence that a
unction.
istance that has been demon-
noter activity. We consider
e because DNA gyrase, the

enzyme in *E. coli* responsible for maintaining the negative superhelical torsion on the DNA double helix, can bind at many sites along the DNA and introduce negative supercoils into the double helix. Because of the covalently continuous structure of the DNA, the energy put into the double helix at the expense of ATP (adenosine triphosphate) hydrolysis can be transmitted at a distance to a promoter site where it may be utilized by the RNA polymerase to unwind the DNA and initiate transcription. This example is rather clear for two reasons: First, transmission of superhelicity is dependent on the covalently closed structure of the DNA that we know to exist in the cell, and second, there is convincing biochemical and genetic evidence that supercoiling matters both *in vivo* and *in vitro* (Gellert 1981). Most of the other proposals for action-at-a-distance fail on one or the other of these two tests for significance.

1.3 STEPS IN TRANSCRIPTION INITIATION

The fundamental motivation for dissecting transcription initiation into separate steps is the observation that transcription initiation occurs at frequencies that range from about one every 5 to 10 sec for the strongest promoters in *E. coli* (i.e., rRNA genes) to frequencies of once per generation (*lacI*) and perhaps even less frequently (Hahn et al. 1977). Thus, if initiation frequency is characterized by a lower limit of once a generation and by an upper bound of once every few seconds, then the dynamic range of initiation frequency and, therefore, promoter function is at least 1,000-fold and perhaps 10,000-fold. The remarkable feature of this entire range of function is that, to a first approximation, only one enzyme in *E. coli* participates in these initiation frequencies. Thus, we can add to the list of reasons for variability within the *E. coli* promoter sequences the notion that initiation frequency is related directly to DNA sequence. It is likely that some maximum may be defined by the consensus sequence and that by decreasing homology to the consensus sequence a decrease in initiation would follow. Clearly, as with any process occurring in living cells, optimum function is not ordinarily equated to maximal function. For the economy of the cell, it is of great advantage to maintain certain promoters initiating at low frequency and others at relatively high frequency.

1.3.1 Pathways to Open Complex Formation

The dissection of RNA chain initiation began many years ago with the work of Zillig and co-workers (Walter et al. 1967) who proposed that there were two principal steps in initiation: (1) a binding of the enzyme followed by (2) the conversion of an inactive intermediate to a complex that was ready to initiate a chain quickly. This scheme was sharpened by the experiments and proposals of Chamberlin (1974) and has now gained wide acceptance as the standard working model for RNA polymerase interaction with the promoters (see von

Hippel et al. 1984; McClure 1985, for reviews). In schematic form, the model can be written as follows:

$$R + P \underset{K_B}{\rightleftharpoons} RP_C \xrightarrow{k_2} RP_O \xrightarrow{NTPs} \rightarrow \cdots \rightarrow \underset{\sigma}{\searrow} \rightarrow \text{RNA}. \qquad (1.1)$$

In this scheme the binding step is characterized by the apparent binding constant, K_B; the conversion from the inactive intermediate (RP_c, closed complex) to the active intermediate (RP_o, open complex) occurs with rate constant, k_2. The binding of triphosphates and elongation from the promoter is assumed to be relatively rapid, although exceptions to this rule are discussed under RNA chain initiation (Section 1.3.2). In the early 1980s, various alterations of this basic scheme have been proposed that do not significantly alter the functional characteristics of this two-step scheme. First is the recognition that K_B, functionally or operationally described, is simply any group of steps involved in the initial binding of the RNA polymerase that results in complexes that are rapidly dissociable. These steps need not be the same for all promoters. The principal characteristic is that the rate of dissociation from this group of complexes is more rapid than k_2. k_2 defines an important step in promoter function inasmuch as the reverse of open complex formation—namely, k_{-2} is usually very slow and is probably not of any physiological significance (lifetimes of open complexes determined *in vitro* are typically on the order of an hour to several hours). What is remarkable is that the conversion from the inactive closed complex to the transcriptionally active open complex is a very slow step indeed. Rate constants measured in several laboratories span a range from 0.1 per sec to 0.001 per sec and even slower (see references in Figure 1–5 legend). Surely the unwinding of DNA discussed earlier in the structural section is involved in this isomerization step. But if that step must occur with all promoters, wherein lies the basis for the variability and the magnitude of k_2? Second, why is the isomerization step so slow compared to other conformational changes seen in other enzymes?

A resolution of the first question has been suggested by work from several laboratories (Kadesch et al. 1982; Roe et al. 1984; Buc and McClure 1984). First, when the isomerization step is studied at different temperatures and at different ionic strengths, it is clear that the rate-limiting step on this pathway is in fact an isomerization but that immediately following that slow conversion from closed to another intermediate, a very rapid conversion occurs under normal physiological conditions. The following reaction scheme incorporates the additional intermediate that is proposed:

$$R + P \underset{K_B}{\rightleftharpoons} RP_C \xrightarrow[\longleftarrow]{k_2} RP_I \xrightarrow{k_3} RP_O. \qquad (1.2)$$

Indeed, the conversion from the second intermediate to RP_o can be observed only at very high salt or at temperatures in the range of 18 to 20°C. The

second question remains
salt dependencies of k_2
formation. From its cha
dependence and relative
major alteration in the DN
entirely. The relative d
follows principally from
fore, it is not easily obse
to the second intermedi
investigation of the puta
what portions of the DN
and what the nature of t

It is relatively easy to
of the transcriptionally a
constants for open comp
combination of the enz
results nearly quantitativ
complex has been investi
and chemical modificatic
open complex has been in
unwound DNA base pai
shown that 17 base pairs
unwinding was qualified
ment was done by deter
RNA polymerase bindir
protection-type experime
11 to 12 base pairs could
exposed to the chemic
extended to the −10 re
number of base pairs uns
that experiment is that
startsite of transcription.
the total perturbation bu
perturbation in the prom

1.3.2 RNA Chain Initiat

The conversion of RNA p
core complex involves
sequence-specific contac
demonstrated originally
eight to nine nucleotides
McClure 1980). The poi
determined. It is not an
of establishing the fact a

schematic form, the model

→ RNA. (1.1)

ed by the apparent binding
intermediate (RP_c, closed
complex) occurs with rate
ngation from the promoter
ns to this rule are discussed
early 1980s, various altera-
it do not significantly alter
me. First is the recognition
s simply any group of steps
se that results in complexes
be the same for all promot-
lissociation from this group
important step in promoter
formation—namely, k_{-2} is
iological significance (life-
ypically on the order of an
it the conversion from the
tive open complex is a very
al laboratories span a range
ee references in Figure 1–5
d earlier in the structural
f that step must occur with
ility and the magnitude of
slow compared to other

gested by work from several
; Buc and McClure 1984).
ferent temperatures and at
iting step on this pathway is
owing that slow conversion
d conversion occurs under
iction scheme incorporates

$\rightleftharpoons RP_O$. (1.2)

ate to RP_o can be observed
range of 18 to 20°C. The

second question remains an enigma. It is not clear from the temperature and salt dependencies of k_2 what kind of step this could be in open complex formation. From its characteristics—namely, the relatively low temperature dependence and relatively low salt dependence—k_2 does not appear to be a major alteration in the DNA structure, although that notion cannot be dismissed entirely. The relative dearth of information about the isomerization step follows principally from the fact that RP_c is a transient intermediate; therefore, it is not easily observable, even though the overall conversion from RP_c to the second intermediate is favorable. What is needed is a more detailed investigation of the putative closed complexes for several promoters to ask what portions of the DNA are contacted by RNA polymerase in that complex and what the nature of those interactions are.

It is relatively easy to study the properties, both structural and functional, of the transcriptionally active open complex. After all, the overall equilibrium constants for open complex formation are in excess of 10^{10} M^{-1}. Therefore, a combination of the enzyme and promoter under physiological conditions results nearly quantitatively in the formation of open complexes. This open complex has been investigated in some detail with the protection experiments and chemical modification experiments described previously. In addition, the open complex has been investigated with respect to the number of topologically unwound DNA base pairs. A report from Gamper and Hearst (1982) has shown that 17 base pairs are unwound in a topological sense. The extent of unwinding was qualified previously to emphasize the fact that the measurement was done by determining the linking number change in DNA due to RNA polymerase binding after exposure to a nicking-closing enzyme. A protection-type experiment performed by Siebenlist (1979) showed that about 11 to 12 base pairs could react with dimethyl sulfate (DMS) and that the region exposed to the chemical reagent was in the vicinity of the startsite and extended to the −10 region. His result is probably a lower estimate for the number of base pairs unstacked and melted. The important fact learned from that experiment is that the DNA melting is localized to a region near the startsite of transcription. The topological experiment provides an estimate of the total perturbation but says nothing about the localization of the structural perturbation in the promoter.

1.3.2 RNA Chain Initiation

The conversion of RNA polymerase from an open complex into a transcribing core complex involves the release of σ and the release of the favorable sequence-specific contacts formed in the open complex. The release of σ was demonstrated originally by Burgess et al. (1969). It was found to occur after eight to nine nucleotides on the synthetic poly [d(AT)] template (Hansen and McClure 1980). The point of σ release on an *E. coli* promoter has not been determined. It is not an easy measurement to make although the importance of establishing the fact and the location of σ release remains an outstanding

and very important problem in promoter function. Although we do not know either of these facts, we can operationally end the initiation phase of RNA synthesis at the point(s) of σ release.

A feature of the RNA polymerase initiation process that may be linked to σ release is the observation of abortive initiation at the promoter. Many promoters, even in the presence of all four triphosphates, will synthesize and release short oligonucleotides that have been properly initiated but not productively elongated into full-length RNA chains (Carpousis and Gralla 1980: Munson and Reznikoff 1981). Even if abortive cycling is a general property of RNA polymerases, it is still not clear why some promoters and not others stimulate a rather prodigious amount of the cycling reaction on the pathway to productive chain elongation. For example, the *lacUV5* and a promoter from Tn5 both initiate abortively with very high frequency *in vitro*, with perhaps 10 to 20 oligonucleotides produced for each full-length chain (Munson and Reznikoff 1981). The consequence of the abortive cycling could be a slow step in overall chain elongation that is independent of the time required to form open complexes as dictated by the combination of K_B and k_2. The addition of the time required to synthesize abortively many oligonucleotides would thus contribute to the overall time required to make a full-length RNA chain. A possible explanation of the abortive initiation reaction from a mechanistic standpoint is that those promoters, which display the cycling reactions, may be those that release the polymerase with some difficulty during translocation out of the promoter.

One could imagine two nonexclusive models for the promoter exit process. The enzyme may bind specifically to the -10 and -35 regions and thread the downstream DNA double helix through the active site while it is transcribing oligonucleotides in the range of 3–6–8. When a stable complex is formed, the specific contacts are released all at once. Alternatively, one could imagine that the DNA is rigid, that the RNA polymerase is flexible, and that the enzyme slides back and forth from the start point to the site(s) at which abortive products are released. In all cases that have been studied, the abortive cycling occurs without the release of RNA polymerase from the promoter, as judged by the resistance of such reactions to competing reagents such as heparin. A question that is still unresolved is whether or not the abortive cycling plays any physiological role. This query goes back to the point raised earlier about the time required for the abortive cycling and its contribution to overall RNA synthesis frequencies. There seems to be little chance of a direct experimental test of the extent of abortive cycling *in vivo* inasmuch as the small oligonucleotides formed *in vivo* would be rapidly degraded by cellular ribonucleases. We suggest instead that some indirect method could be devised in which a correlation between *in vivo* and *in vitro* properties might be improved as a result of incorporating into the model for RNA synthesis the time required for the abortive cycling. This would be a particularly attractive possibility if there were mutations that affected the frequency of abortive cycling for a particular promoter. One candidate is a mutant derivative of the P22 P_{sar} promoter (S.-M. Liao and W. R. McClure, 1985, unpublished).

Finally, the abortive
exploited originally by M
assays for promoter streng
and k_2 to overall initiation
A summary of promoter
Figure 1–5 in the form of
values for K_B and k_2 for p
well as weak. It is clear
roughly three orders of
picture is that K_B and k_2 i
of RNA chain initiation. I
1982), the strongest pron
corner where they are ch
k_2; that is, those promoter
the inactive intermediate i

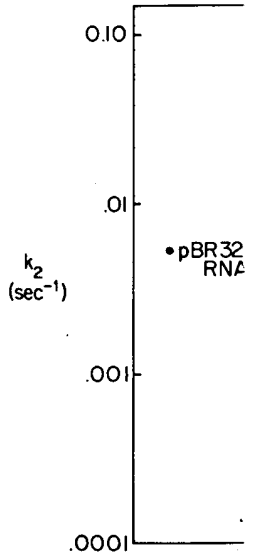


FIGURE 1–5 A selectivity
placed on the selectivity ma
were determined *in vitro*. T
both scaled logarithmically
promoter strength measuren
McClure 1980); *lac* (Malan
(Hawley and McClure 1982;
and Gussin 1983b); T7 A3 a
et al. 1984); p^{IN} and p^{OUT} (Si
sons between wild-type and

)n. Although we do not know
the initiation phase of RNA

process that may be linked to
ıtion at the promoter. Many
ıosphates, will synthesize and
ıroperly initiated but not pro-
; (Carpousis and Gralla 1980;
ycling is a general property of
ne promoters and not others
ing reaction on the pathway to
lacUV5 and a promoter from
ıency *in vitro*, with perhaps 10
ll-length chain (Munson and
ve cycling could be a slow step
: of the time required to form
ı of K_B and k_2. The addition of
ıy oligonucleotides would thus
:e a full-length RNA chain. A
. reaction from a mechanistic
ılay the cycling reactions, may
difficulty during translocation

els for the promoter exit proc-
10 and −35 regions and thread
active site while it is transcrib-
en a stable complex is formed,
lternatively, one could imagine
ɔrase is flexible, and that the
point to the site(s) at which
that have been studied, the
of RNA polymerase from the
eactions to competing reagents
esolved is whether or not the
his query goes back to the point
bortive cycling and its contribu-
re seems to be little chance of a
:ive cycling *in vivo* inasmuch as
vould be rapidly degraded by
some indirect method could be
:o and *in vitro* properties might
he model for RNA synthesis the
ould be a particularly attractive
:ted the frequency of abortive
ate is a mutant derivative of the
lure, 1985, unpublished).

Finally, the abortive cycling feature of RNA polymerase initiation was exploited originally by McClure and co-workers and other laboratories in assays for promoter strength, which separate *in vitro* the contributions of K_B and k_2 to overall initiation frequency (see references in legend to Figure 1–5). A summary of promoter strength data from several laboratories is shown in Figure 1–5 in the form of a selectivity map—that is, a compilation of several values for K_B and k_2 for promoters that have been characterized as strong as well as weak. It is clear that the spectrum of both K_B and k_2 ranges over roughly three orders of magnitude. The hypothesis that follows from this picture is that K_B and k_2 in combination can determine significantly the rates of RNA chain initiation. In the selectivity map representation (Hawley et al. 1982), the strongest promoters occupy the region in the upper right-hand corner where they are characterized by high values of K_B and high values of k_2; that is, those promoters have a very high affinity for RNA polymerase, and the inactive intermediate is converted rapidly into the transcriptionally active

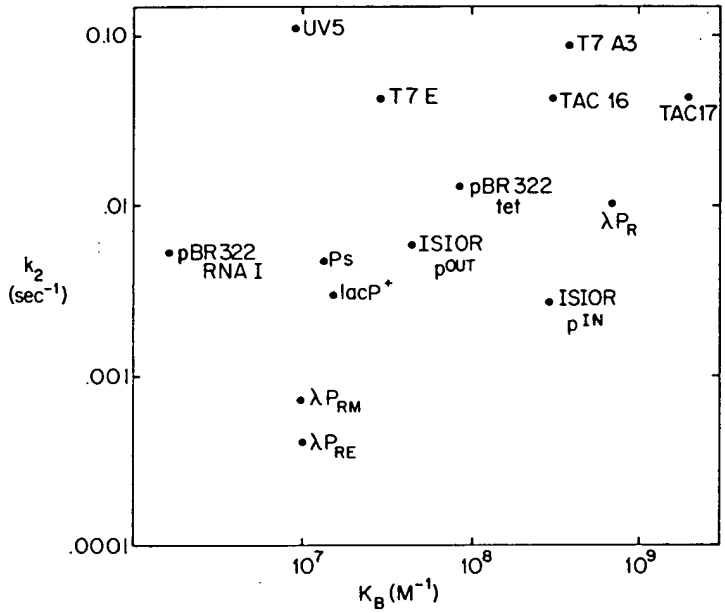


FIGURE 1–5 A selectivity map for several *E. coli* promoters. Each promoter was placed on the selectivity map according to the reported values for K_B and k_2, which were determined *in vitro*. The ordinate (k_2 values) and the abscissa (K_B values) are both scaled logarithmically to accommodate the range of values. References for the promoter strength measurements are: TAC (Mulligan et al. 1985); λP_R (Hawley and McClure 1980); *lac* (Malan et al. 1984); RNA I (Wood and Lebowitz 1984); λP_{RM} (Hawley and McClure 1982; Shih and Gussin 1983a); λP_{RE} (Shih and Gussin 1984; Shih and Gussin 1983b); T7 A3 and T7E (Prosen and Cech 1985); *tet* (Bertrand-Burggraf et al. 1984); p^{IN} and p^{OUT} (Simons et al. 1983). References to earlier work and comparisons between wild-type and mutant promoters are also cited in these publications.

complex. At the opposite end of the spectrum, the weak promoters have rather low affinities for RNA polymerase and are only slowly converted into the transcriptionally active open complex. At least numerically, this compilation of promoter strengths can be seen to be in agreement with the range of initiation frequencies discussed at the beginning of the section; namely, the times on the order of 10 sec or less for the strongest promoters and an hour or more for the weakest promoters are going to be characteristic of the frequency of RNA chain initiation from those promoters. An issue still to be resolved is the relative contribution of the concentration and/or activity of RNA polymerase within the cell, and that is a topic with which we deal in more detail in Section 1.4.

1.4 STRUCTURE-FUNCTION CORRELATION

In this section we treat three related topics separately and indicate at the end the interrelationship of all three. Basically, we relate the structure of the promoter discussed in Section 1.3 to the functional steps that have also been described. Then we consider the relevance of the *in vitro* investigation of promoter strength to *in vivo* gene expression, and finally we consider more complicated cases of promoter function in which closely spaced promoters can have an effect on one another and, therefore, on gene expression.

1.4.1 The Bipartite Model

One of the first suggestions for promoter structure-function followed the proposal that the promoter had a bipartite structure—that is, a -35 and a -10 region. Quite logically, then, Gilbert (1976) suggested that the -10 region was involved in the melting-in of the RNA polymerase near the startsite and that the -35 region participated in the recognition of RNA polymerase for the promoter. The bipartite model ran into some modest difficulty with some of the early measurements of the effect of mutations on promoter function. In particular, Hawley and McClure (1980) showed that a -35 region mutation (in the λP_R *x3* mutant) had an effect on both K_B and k_2. At about the same time, Stefano and Gralla (1982b) were characterizing -35 region mutations in the *lac* promoters and found that they affected only k_2. Hawley and McClure (1982) also showed that an up-promoter mutation in the λP_{RM} promoter increased both K_B and k_2. At that time, two hypotheses were proposed to deal with the experimental findings. Hawley and McClure chose to preserve the bipartite model and indicated how a single base pair change in the -35 region could affect only one step in open complex formation although it appears to affect both K_B and k_2. The other idea, proposed by Stefano and Gralla, states that there is simultaneous recognition at the -10, the -35, and the spacer element of the promoter when RNA polymerase was forming the closed complex. The two ideas are really only extremes of a more general model that says it is promoter but that multiple p tion initiation (see Chapter between the two extremes or

A further mechanistic that as indicated, K_B corres rapid with respect to the complex. It is certainly not included in the rapid equili even if the overall pathwa promoters (by no means es model have also been obta phage P22 (McClure et al. are at least two mutations in there is one mutation in the be classed as successful cl model, although the compl sion of these results as w information on the putative

1.4.2 Promoter Strength

A rather attractive directio and promoter strength is th DNA sequence alone. A precisely along those line algorithm for evaluating homology score, *in vitro* p as $K_B \times k_2$ correlated wel The correlation was linea deviation of ± 4 within th schemes is where to get the Staden (1984) used essen compiled by Hawley and homology score, using ratl the contribution of each ba Mulligan shows that the St correlation as the Mulliga ments for promoter evalua to promoter function is ad ate or even long-range c principally affect the con strength. This assumptio enough measurements of strength and selectivity.

general model that says it is likely that a preferred pathway exists for a given promoter but that multiple pathways are possible. This is also true for translation initiation (see Chapter 7). Unfortunately, evidence that could distinguish between the two extremes or even the intermediate view has not been obtained.

A further mechanistic complexity that should be appreciated is the fact that as indicated, K_B corresponds to a group of equilibrium constants that is rapid with respect to the conversion of the closed complex to the open complex. It is certainly not necessary that the same group of complexes be included in the rapid equilibrium phase of the reaction for each promoter, even if the overall pathway to open complex formation is identical for all promoters (by no means established). Additional challenges to the bipartite model have also been obtained in a study of the P_{ant} promoter of bacteriophage P22 (McClure et al. 1983; and unpublished). For that promoter there are at least two mutations in the -35 region that affect only k_2, and conversely, there is one mutation in the -10 region that affects only K_B. These data must be classed as successful challenges to the simplest version of the bipartite model, although the complexities discussed earlier are apropos in the discussion of these results as well, particularly in the absence of any structural information on the putative intermediates involved in this reaction.

1.4.2 Promoter Strength Prediction

A rather attractive direction with respect to the relationship of DNA sequence and promoter strength is the possibility of predicting promoter strength using DNA sequence alone. A proposal by Mulligan et al. (1984) was designed precisely along those lines, and in fact, they claim that with a particular algorithm for evaluating DNA sequence and converting it to a so-called homology score, *in vitro* promoter selectivity by RNA polymerase expressed as $K_B \times k_2$ correlated well with the DNA sequence-based homology score. The correlation was linear over four orders of magnitude, with a standard deviation of ± 4 within the entire interval. The dilemma in such predictive schemes is where to get the score for each promoter of a given DNA sequence. Staden (1984) used essentially the same database [i.e., the 112 promoters compiled by Hawley and McClure (1983a)] to make the equivalent of a homology score, using rather different assumptions about the partitioning of the contribution of each base pair to the overall score. Unpublished work from Mulligan shows that the Staden algorithm gives essentially the same degree of correlation as the Mulligan algorithm. The assumption in both these treatments for promoter evaluation is that the contribution of individual base pairs to promoter function is additive (i.e., additivity of ΔGs) and that the immediate or even long-range context of the DNA for each promoter does not principally affect the contribution of an individual base pair to promoter strength. This assumption is so far untested, largely because there are not enough measurements of the individual effects of base pairs on promoter strength and selectivity.

Moreover, the contribution of spacer length to promoter function may not be identical for all classes of promoters. It has been found, for example, that the spacer length (16, 17, and 18) for the semisynthetic TAC promoter (Mulligan et al. 1985) is less important than for the weaker *lac* P^s promoters characterized by Stefano and Gralla (1982a). Thus, there seems to be room for many contributions to the area of predicting promoter strength. An interim goal along this line is the automated use of the evaluation algorithms for the prediction of promoter sites within DNA sequence files. The program described by Mulligan et al. (1984) has successfully predicted the location of three previously unknown promoters, and that is encouraging. The general applicability and success of these algorithms is expected to increase in the future, making it possible to locate promoters with some confidence in an objective way that will save biochemists some experimental time.

1.4.3 Correlation of *in Vitro* Promoter Strength to *in Vivo* Expression

In the meantime, we can examine the evidence that the DNA sequence and *in vitro* characterization of promoters does correspond to *in vivo* function. The evidence comes from the following several lines of experimentation. First, down-promoter mutations known to reduce expression *in vivo* are without exception down *in vitro* with respect either to measured values of K_B or k_2 or both. Likewise, the up-promoter mutations selected and characterized *in vivo* are always up *in vitro*. Second, activators enhance one or the other or both of the two steps in open complex formation (Hawley and McClure 1982; Malan et al. 1984; Shih and Gussin 1983b), and repressors have been shown to inhibit one or the other of the two steps in initiation (Hawley et al. 1985). Finally, the effect of supercoiling on promoters *in vitro* corresponds rather closely with the effect of supercoiling *in vivo*, as noted in Section 1.2. The effects of supercoiling *in vivo* are currently under active investigation in several laboratories, and the results show that supercoiling, although contributing energetically in a favorable sense to open complex formation, has been found to increase and also to decrease the expression of certain promoters (Gellert 1981). The results obtained *in vivo*, while intriguing, are subject to certain caveats; in particular, even if an effect of supercoiling can be shown to be immediate and specific to a particular promoter using drugs that interfere with gyrase activity or gyrase temperature-sensitive mutants, it is still possible that the overall effect on expression could be an indirect effect resulting from the binding or release of a control protein near the promoter. The alternate possibility that supercoiling has a direct effect on initiation frequency can only be tested *in vitro* using purified systems. For example, Malan et al. (1984) have shown that supercoiling is required *in vitro* to see high promoter strength at the *lac* P^+ promoter in agreement with the earlier findings of Sanzey (1979), who performed her experiments *in vivo*. An example in which supercoiling plays a regulatory role is discussed later.

A correlation betwe
relation to *in vivo* expre
in vivo data obtained witl
well as *lacZ* fusions to va
represented in that study
between *in vitro* and *in*
originated with that study
RNA polymerase active c
served for a concentrati
corresponds to the *in vit*
ments were made. The su
centration *in vivo* is in the

1.4.4 Contributions fror

Finally, in this section we
interact in three differen
promoters such as the *rr*
message is transcribed fr
to different conditions or
lute level of the mRNA t
where transcription occu
tions for transcribing diffe
in which a common contr
between the start points
repress one promoter whi
gent geometry and contro
the lamdoid phages).

A third class is mor
interesting. The converge
over a common region of
only one strand actually
template for an antisense
the message. In some ca
polymerases has been est
1979; Schmeissner et al. 1
close to one another and,
exclusive binding to one
two promoters are separ
distance dependence was
used constructions involvi
promoter of λ reading the

The interactions betw
ined thus far have been c
sive binding. It is also

A correlation between *in vitro* measurements of K_B and k_2 and their relation to *in vivo* expression was suggested by McClure (1983) based on *in vivo* data obtained with the *lac* promoter and *lac* promoter mutations, as well as *lacZ* fusions to various λ promoters. The range of promoter strength represented in that study was only about one hundredfold; the agreement between *in vitro* and *in vivo* was relatively good. An interesting hypothesis originated with that study because the correlation required the setting of an RNA polymerase active concentration *in vivo*. The best agreement was observed for a concentration of active polymerase *in vivo* of 30 nM if that corresponds to the *in vitro* conditions under which the K_B and k_2 measurements were made. The suggestion that the RNA polymerase activity or concentration *in vivo* is in the range of 30 nM is untested and remains speculative.

1.4.4 Contributions from Promoter Geometries

Finally, in this section we consider how closely spaced *E. coli* promoters can interact in three different geometries. Rather well known are the tandem promoters such as the *rrn*, *gal*, and T7 early promoters wherein the same message is transcribed from tandem promoters that can in principle respond to different conditions or levels of cellular constituents to regulate the absolute level of the mRNA transcript. A second category is the divergent class where transcription occurs from closely spaced promoters in opposite directions for transcribing different messages. This results in a situation for control in which a common control molecule or ancillary protein can bind to a region between the start points of transcription and, for example, simultaneously repress one promoter while activating another. Many examples of this divergent geometry and control property are known (e.g., the P_{RM} -P_R complex of the lamdoid phages).

A third class is more complicated and, in part, for that reason more interesting. The convergent class where two promoters oppose one another over a common region of DNA is transcribed in both directions. In all cases, only one strand actually contains the sense message; the other strand is a template for an antisense RNA that could in principle control expression of the message. In some cases, transcriptional interference between the two polymerases has been established by *in vivo* experiments (Ward and Murray 1979; Schmeissner et al. 1980). In other cases, the two promoters can be very close to one another and, as a consequence, interfere because of the mutually exclusive binding to one or the other of the two promoters. The farther the two promoters are separated, the less interference will be observed. The distance dependence was first suggested by Ward and Murray (1979), who used constructions involving the *trp* promoter of *E. coli* in opposition to the P_L promoter of λ reading the *lacZ* gene.

The interactions between closely spaced promoters that have been examined thus far have been characterized by interference due to mutually exclusive binding. It is also formally possible that closely spaced polymerase

molecules in any of the orientations discussed could result in enhancement of transcription initiation. Exactly that claim has been made by Travers et al. (1983) for transcription of the tRNA tyrosine gene in *E. coli*. Other documented cases have not been described.

Finally, in the category of other roles for polymerase, the enzyme could, as viewed earlier, be imagined as a repressor by binding to a site that controls the function of other RNA polymerases at adjacent sites (a possible example of this situation is described later). These and other roles for RNA polymerase in complex geometries may suggest a regulatory role for the enzyme, in addition to the regulation already imposed on the initiation of RNA synthesis. It will not be surprising to find that an abundant DNA binding protein, such as RNA polymerase, can function as a *trans* regulator as well as an important enzyme in the cell.

1.5 REGULATION OF TRANSCRIPTION INITIATION

One of the early critical physiological observations in the field of molecular biology was that the expression of many genes is regulated. In particular, many genes that encode catabolic functions are expressed only when the cells are presented with the relevant pathway's substrate, and many genes that encode anabolic functions are expressed at lower levels when the cells are presented with the relevant pathway's end product. The early studies at the Pasteur Institute on the *E. coli lac* and *trp* operons led to the conclusion that bacterial gene regulation is often negative in character (the active regulatory macromolecule turns off gene expression) (Jacob and Monod 1961; Cohen and Jacob 1959).

Soon after the negative regulation model for gene expression control was proposed, studies on the *ara* operon by Englesberg and his colleagues (1965)

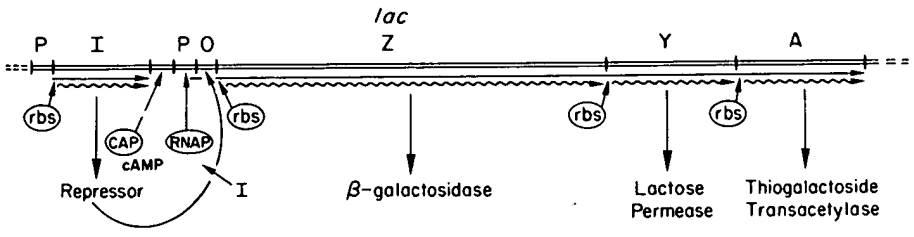


FIGURE 1-6 The *lac* operon. The transcription of the *lacZ*, *Y*, and *A* genes results from transcription initiation by RNA polymerase (RNAP) at *lacP*. This event is regulated in a positive fashion by the catabolite gene activator protein (CAP) when it is complexed with cAMP, and in a negative fashion by the *lacI* gene product, the repressor. The repressor in turn is inactivated by the inducer (*I*). Translation (indicated by the wavy arrows) of each of the product proteins results from ribosome (rbs) binding to the mRNA translation initiation signal for each gene. Reproduced with permission from Reznikoff (1984).

suggested that positive cc
large number of systems
ized catabolite repressic
catabolic operons is an e
1970; Emmer et al. 1970)
now known to show du
negative or positive chara
protein binds relative to
the CAP protein vis-à-vis
1977; Adhya and Miller
apparently functioning in
of the enzyme, RNA poly
structure of the promote
changes in the orientatio

1.5.1 Negative Regulat

The *lac* operon regulato
components: the *lac* repr
the *lac* operator (*O*) (Fig
repressor binds, thereby
complex with *lacP*; an
allolactose that, when b
the repressor's configur
(reviewed in Jacob and
and Reznikoff 1978). The
other systems, and thus.
The galactose operon i
variation.

The repressor was d
the *lacI* gene give rise
mutations (designated *I*
low levels of *lac* express
tions that specifically d
terminal portion of the c
a few defined internal
1978). Biochemical stuc
(Beyreuther 1978; Webe
sequence comparisons h
and some other known
operator binding sites (F

The operator is defi
O^c) that are located with
(see Figure 1-1). This
repressor binding (filter
O^c mutations reduce thi

suggested that positive control of gene expression might also occur. In fact, a large number of systems are under positive control. For example, the generalized catabolite repression system that controls the *lac* and several other catabolic operons is an example of positive control (Schwartz and Beckwith 1970; Emmer et al. 1970). In addition, several genetic regulatory systems are now known to show dual control effected by the same protein, with the negative or positive character of the regulation being dependent on where the protein binds relative to the promoter (i.e., the λcI repressor vis-à-vis λP_{RM}, the CAP protein vis-à-vis the *gal* P_{G1} and P_{G2}) (Meyer et al. 1975; Musso et al. 1977; Adhya and Miller 1979). Finally, other modes of gene regulation are apparently functioning in some systems, such as modifications in the structure of the enzyme, RNA polymerase (Grossman et al. 1984), modifications in the structure of the promoter (Kleckner et al. 1984; Yin and Reznikoff 1986), and changes in the orientation of the promoter (Zieg et al. 1977).

1.5.1 Negative Regulation of Transcription Initiation

The *lac* operon regulatory system is illustrated in Figure 1–6. It involves three components: the *lac* repressor, a tetrameric protein encoded by the *lacI* gene; the *lac* operator (*O*) (Figure 1–1), the target DNA sequence to which the *lac* repressor binds, thereby preventing RNA polymerase from forming an open complex with *lacP*; and the inducer, β-galactoside, compounds such as allolactose that, when bound to the repressor, cause an allosteric change in the repressor's configuration, which reduces its affinity for the operator (reviewed in Jacob and Monod 1961; Beckwith and Zipser 1970; and Miller and Reznikoff 1978). The general features of the *lac* operon are similar to most other systems, and thus, it is used as a model to discuss negative regulation. The galactose operon is discussed briefly since it presents an interesting variation.

The repressor was defined initially by genetic analyses. Null mutations in the *lacI* gene give rise to a constitutive phenotype. Some specific point mutations (designated I^s) make the operon uninducible (e.g., programming low levels of *lac* expression in the presence of the inducer). Missense mutations that specifically define the operator binding site are located in the N terminal portion of the coding sequence, while the I^s mutations are located at a few defined internal locations that form the inducer binding site (Miller 1978). Biochemical studies have further defined the repressor's properties (Beyreuther 1978; Weber and Geisler 1978; Barkley and Bourgeois 1978), and sequence comparisons have suggested to some authors that the *lac* repressor and some other known repressors have a common structural feature in their operator binding sites (Pabo and Sauer 1984).

The operator is defined by *cis*-dominant constitutive mutations (designated O^c) that are located within a symmetry element in the *lac* regulatory sequence (see Figure 1–1). This sequence has been shown to participate directly in repressor binding (filter binding studies and protection experiments), and the O^c mutations reduce this binding (Barkley and Bourgeois 1978). As with most

other operators (an exception is described in the paragraph below), this symmetry element is located in a position relative to the promoter that would be consistent with a mutually exclusive binding mechanism; that is, when the repressor is bound, RNA polymerase cannot bind. Some *in vitro* experiments, such as that reported by Majors (1975), are consistent with such a model. Thus, the general mechanism for how most repressors act is probably determined.

At least one system, the *gal* operon, may function quite differently. The critical factor is that the *gal* operator has a unique structure for two reasons:

1. The *gal* operator has two components: one on the regulatory element region and one downstream in the first structural gene (Irani et al. 1983). O^c mutations have been isolated in both sites, suggesting that *gal* repressor subunits must bind (simultaneously?) at both sites to achieve complete repression. [There is a hint of this type of arrangement in *lac* since a secondary *lac* repressor binding site is located in the first structural gene, but for *lac* this binding site may not be critical for the repression process (Reznikoff et al. 1974; Gilbert et al. 1975).]
2. The *gal* operator component in the controlling elements is located outside the presumed promoter sequence; at positions -67 to -53 (for P_{G1}) and -62 to -48 (for P_{G2}) (Adhya and Miller 1979; DiLauro et al. 1979). Thus, its location does not lend itself to a simple repressor-operator/RNA polymerase promoter competitive binding model. The mechanism of the *gal* repressor action is unknown but may involve some sort of action-at-a-distance as discussed before.

It would be convenient in the construction of regulated expression vehicles if repressors could act efficiently by blocking transcription elongation. This would allow one to choose a strong promoter and merely place the operator at any convenient location downstream. Unfortunately, experiments designed to test the efficacy of this type of regulation have shown that repressors are significantly less efficient at blocking elongation than they are at blocking initiation. For instance, the *lac* repressor could only regulate readthrough transcription three- to tenfold in *trp-lac* fusions when acting at the operator, and it was totally ineffective at blocking readthrough transcription when acting at the secondary repressor binding site (Reznikoff et al. 1969). The *trp* repressor was found to be totally ineffective at blocking readthrough transcription in λP_L-*trp* fusion constructs (Franklin 1971).

1.5.2 Positive Regulation of Transcription Initiation

The critical genetic criteria in defining a positive control system are (1) the isolation of mutations that fail to express the system in question because they fail to synthesize the positive regulatory protein (defined by criteria such as

genetic mapping experi
nonsense mutations, an
mutations that alter a *cis*-
1–6 for a description c
expression). The identifi
esting complications, wh

As discussed befor
the target site for a positi
RNA polymerase. Super
the maximal level of tra
logical criterion for mak
and his colleagues (Beck

Lac promoter (RNA
responsive to the presen
CAP-cAMP complex), v
should have dramatically
should have the same le
wild-type sequence. This
In some cases [such as
positive activator may ov
case a sequence change
sequence. In addition, s
binding site, although qu
not manifest the corre
reflecting the mechanis
scription. The direct tes
biochemical (although h
could perform protectio
tein is bound on the DN
wild-type sequences. Sc
and biochemical criteria
next paragraph, mutati
criteria but that are loc
determined by protectio

The unusual mutati
alter the spacing betwe
moter (Caruthers et al. 1
Mandecki and Caruther
effect on CAP-cAMP s
region (either the spe
mechanism(s) of this p
CAP-cAMP binding site
(Queen and Rosenberg
deCrombrugghe et al. 1

genetic mapping experiments, complementation studies, the existence of nonsense mutations, and the sequence alterations) and (2) the isolation of mutations that alter a *cis*-acting target site for the activator protein (see Figure 1–6 for a description of the positive regulatory elements that affect *lac* expression). The identification of the target site in particular has some interesting complications, which are discussed later.

As discussed before, it is critical, although difficult, to distinguish the target site for a positive activator protein from the sequence recognized by RNA polymerase. Superficially, they both have a similar role: determining the maximal level of transcription initiation. The qualitative genetic/physiological criterion for making this distinction has been developed by Beckwith and his colleagues (Beckwith et al. 1972; Hopkins 1974; Beckwith 1981).

Lac$^-$ promoter (RNA polymerase binding site) mutations should still be responsive to the presence or absence of the positive activator protein (the CAP-cAMP complex), while *Lac*$^-$ mutations in the CAP-cAMP binding site should have dramatically reduced sensitivity to the CAP-cAMP complex and should have the same level of expression in the absence of CAP-cAMP as the wild-type sequence. This is a powerful approach, but it has several limitations. In some cases [such as *gal* (Taniguchi et al. 1979)], the binding site for the positive activator may overlap with the RNA polymerase binding site, in which case a sequence change in one binding sequence is likely to perturb the other sequence. In addition, some mutations that are in the *lac* RNA polymerase binding site, although qualitatively responsive to the CAP-cAMP complex, do not manifest the correct quantitative response (Hopkins 1974), perhaps reflecting the mechanism(s) of CAP-cAMP-mediated stimulation of *lac* transcription. The direct test for the sequence location of the two binding sites is biochemical (although here, too, there are complications). For example, one could perform protection experiments to determine where the relevant protein is bound on the DNA or perform binding studies comparing mutant and wild-type sequences. Some mutations fulfill both the genetic/physiological and biochemical criteria for defining the *lac* CAP site, but as described in the next paragraph, mutations also exist that fulfill the genetic/physiological criteria but that are located outside the *lac* CAP binding site sequence, as determined by protection experiments (Figure 1–7).

The unusual mutations mentioned here are deletions or insertions that alter the spacing between the *lac* CAP-cAMP binding site and the *lac* promoter (Caruthers et al. 1982; LeClerk and Istock 1982; Yu and Reznikoff 1984; Mandecki and Caruthers 1984). Some of these are shown in Figure 1–7. Their effect on CAP-cAMP stimulation of *lac* expression must indicate that this region (either the specific sequence or the spacing) is critical to the mechanism(s) of this process. This finding presents a paradox since the CAP-cAMP binding site is located at different positions in different systems (Queen and Rosenberg 1981; McKay and Steitz 1981; O'Neill et al. 1981; deCrombrugghe et al. 1984).

FIGURE 1-7 Mutations that alter the CAP-cAMP stimulation of *lacP*. The sequence presented and general landmarks are the same as in Figure 1-1. Chemical probe studies (Majors 1977) and comparative sequence analyses (deCrombrugghe et al. 1984) have indicated that the *lac* CAP site is located from −72 to −53. Deletions ending between −72 and −71, −71 and −70, −66 and −65, −65 and −64, and −64 and −61 and −60 and the point mutations between −68 and −57 also define the CAP binding site (these mutations have previously been described by Dickson et al. 1977; LeClerc and Istock 1982; Yu and Reznikoff 1984; Kunkle 1984). There are two other types of mutations that alter CAP-cAMP's ability to stimulate *lac* expression that appear to be unrelated to the sequence-specific binding of CAP-cAMP and, therefore, must affect the mode of action of bound CAP-cAMP. These mutations are the deletions and insertions between −45 and −36 (these affect the CAP site–*lacP* spacing, as well as the P2 −35, −10 spacing) and single bp changes at −34 and −32 (these affect the −10 region of P2 and the −35 region of *lacP*) (LeClerc and Istock 1982; Reznikoff 1984; Yu and Reznikoff 1984; Kunkle 1984; Peterson and Reznikoff 1984a).

Unlike negative regu
tive regulation can be effe

1. The positive activato
directly with RNA
with the promoter (
studies suggest that
activates RNA polyı
gene that result in a
have no apparent eff
have been described
and McClure 1983b
quirements of the *la*
2. The positive activato
the promoter (Dicks
that CAP binding sit
regions in different s
DNA bending of the
and Crothers 1984).
ties of the CAP site
altered by these mut
the DNA conformati
3. The positive activat
would be of an indire
binding of RNA pol
promoter, competes
promoter (McClure
McClure 1984; Mala
is a more extreme e
interact (see discussi
are not mutually exc
could be acting in ar

1.5.3 DNA Supercoilin

As mentioned, supercoil
tion process for many p
increases and in some c
complete surprise when
sion of the very genes wh
(*gyrA* and *gyrB*) are reg
templates were found to
specific inhibitors of D
expression.

Unlike negative regulatory systems, it is not intuitively obvious how positive regulation can be effected. Three possible mechanisms have been proposed:

1. The positive activator may stimulate transcription initiation by interacting directly with RNA polymerase and somehow enhancing its interaction with the promoter (Gilbert 1976). Genetic evidence and model-building studies suggest that this is the mechanism by which the λcI protein activates RNA polymerase interaction with λP_{RM}. Mutations in the λcI gene that result in a failure of the λcI protein to activate λP_{RM} but that have no apparent effect on the binding of the protein to its DNA target site have been described (Guarente et al. 1982; Hochschild et al. 1983; Hawley and McClure 1983b). This model would explain the precise spacing requirements of the *lac* system.
2. The positive activator acts by perturbing the configuration of the DNA at the promoter (Dickson et al. 1975). This model is compatible with the fact that CAP binding sites are located at different positions vis-à-vis promoter regions in different systems and may be related to the CAP-cAMP-induced DNA bending of the *lac* control elements *in vitro* (Kolb et al. 1983; Wu and Crothers 1984). It does not appear to be compatible with the properties of the CAP site promoter spacer mutations in *lac* unless the sequence altered by these mutations plays some role in generating or transmitting the DNA conformational change induced by CAP-cAMP binding.
3. The positive activator acts as a repressor. In this case, the activation would be of an indirect nature. The activator could repress nonproductive binding of RNA polymerase at a site that, because it overlaps with the promoter, competes with productive binding of RNA polymerase at the promoter (McClure et al. 1982; Peterson and Reznikoff 1984a; Malan and McClure 1984; Malan et al. 1984). The *lac* P2 site may be such a site. This is a more extreme example of how closely spaced *E. coli* promoters can interact (see discussion in Section 1.4.4). The three proposed mechanisms are not mutually exclusive; thus, it would be possible that more than one could be acting in any given system.

1.5.3 DNA Supercoiling as a Regulatory Mechanism

As mentioned, supercoiling has been found to affect the transcription initiation process for many promoters (Gellert 1981). In some cases, supercoiling increases and in some cases it decreases gene expression. Thus, it was not a complete surprise when Menzel and Gellert (1983) reported that the expression of the very genes whose products are responsible for generating supercoils (*gyrA* and *gyrB*) are regulated by the level of supercoiling. *In vitro*, relaxed templates were found to program more *gyrB* product synthesis, and *in vivo*, specific inhibitors of DNA gyrase were found to enhance *gyrA* and *gyrB* expression.

FIGURE 1-7 Mutations that alter the CAP-cAMP stimulation of *lacP*. The sequence presented and general landmarks are the same as in Figure 1-1. Chemical probe studies (Majors 1977) and comparative sequence analyses (deCrombrugghe et al. 1984) have indicated that the *lac* CAP site is located from −72 to −53. Deletions ending between −72 and −71, −71 and −70, −66 and −65, −65 and −64, and −61 and −60 and the point mutations between −68 and −57 also define the CAP binding site (these mutations have previously been described by Dickson et al. 1977; LeClerc and Istock 1982; Yu and Reznikoff 1984; Kunkle 1984). There are two other types of mutations that alter CAP-cAMP's ability to stimulate *lac* expression that appear to be unrelated to the sequence-specific binding of CAP-cAMP and, therefore, must affect the mode of action of bound CAP-cAMP. These mutations are the deletions and insertions between −45 and −36 (these affect the CAP site–*lacP* spacing, as well as the P2 −35, −10 spacing) and single bp changes at −34 and −32 (these affect the −10 region of P2 and the −35 region of *lacP*) (LeClerc and Istock 1982; Reznikoff 1984; Yu and Reznikoff 1984; Kunkle 1984; Peterson and Reznikoff 1984a).

1.5.4 Changing Transcription Initiation Specificity by Changing the RNA Polymerase Sigma Subunit

RNA polymerase from *E. coli* has the following subunit structure: β,β',α_2 and σ. β,β',α_2 is the core enzyme that catalyzes mRNA elongation, but it lacks initiation specificity. When the critical relationship between the σ subunit and the transcription initiation process was realized, it was postulated that one possible mode of gene regulation involved the recognition of different classes of promoters by holoenzymes containing different types of σ subunits (Burgess et al. 1969). Although this mode of regulation was not known to exist in *E. coli* until recently, the developmental gene regulation that occurs during *B. subtilis* sporulation and subsequent to phage infection of *B. subtilis* has been known to be correlated with the substitution of different σ subunits for that subunit that dominates the holoenzyme during logarithmic growth (Losick and Pero 1981). Recent experiments from Gross's laboratory indicate that in at least one case, such gene regulation through σ subunit substitution occurs in *E. coli* as well. The positive regulator of the heat shock response in *E. coli* (the product of the *htpR* gene) is a new σ subunit for RNA polymerase (Grossman et al. 1984); that is, mixing purified *htpR* product with core RNA polymerase generates an RNA polymerase that specifically initiates transcription at heat shock promoters. We have no information as to whether additional growth conditions like anaerobiosis result in the utilization of other (different) σ subunits, but the possibility that this type of transcriptional regulation exists for other *E. coli* systems should now be considered.

1.5.5 DNA Methylation: Another Mechanism for Altering the Frequency of Transcription Initiation

Transcription initiation involves the precise recognition of the promoter DNA sequence; thus, it might not be surprising if modification of bases within this sequence were found to lead to an altered frequency of transcription initiation. One common type of base modification is the formation of N^6-methyl adenine in the sequence GATC, which is catalyzed by the *E. coli dam* adenine methylase. Experiments have shown that the transposase mRNAs encoded by Tn5 (Yin and Reznikoff 1985), Tn10, and Tn903 (Kleckner et al. 1984) are synthesized more frequently when the template is fully unmethylated at *dam* sites than in the fully methylated state. An inspection of the relevant promoter sequences suggests an explanation for this observation. The Tn5 transposase promoter contains the sequence

$$\underset{\underline{\text{GATC}}\text{T}\underset{}{\underline{\text{GATC}}}\text{TTCCA},}{\overset{-10 \qquad +1}{\phantom{\text{GATCTGATCTTCCA}}}}$$

in which two *dam* sites overlap the -10 region. The Tn10 transposase promoter sequence is

$$\underset{\underline{\text{GATC}}\text{AAATGGTTTCG}.}{\overset{-10 \qquad +1}{\phantom{\text{GATCAAATGGTTTCG}}}}$$

It is obvious that recogni
dam methylation. If one
more active at programm
could be seen as a mec
transposition) in respons
replicating fork, repair s
the sequence.

1.6 CONCLUSION

The transcription initiati
biochemical studies (lar
analyses. Although there
of studies have allowed u
dence and to obtain an
This type of information
synthetic promoter (Sob
steps involved in the tra
about how these steps ar
Some of the mechanism
have been described. Th
repressor-operator inter
effectively to construct r
of positive control syste
been determined. Oth
discovered. In conclusi
initiation process, but th

REFERENCES

Adhya, S., and Miller, W.
Aiba, H., Adhya, S., and d
An, G., and Friesen, J.D.
Arditti, R.R., Grodzicker,
Barkley, M., and Bourgeoi
eds. Cold Spring
pp. 177–200.
Beckwith, J. (1981) *Cell* 2
Beckwith, J., Grodzicker,
Beckwith, J., and Zipser,
Laboratory, Cold Spr
Bertrand, K.P., Postle,
Gene 23:149–156.
Bertrand, K.P., Postle, K
158:910–919.
Bertrand-Burggraf, E., I
12:1697–1706.

It is obvious that recognition of these −10 regions could be affected by the *dam* methylation. If one assumes that the hemimethylated promoters are also more active at programming transcription initiation, then this type of control could be seen as a mechanism for stimulating transposase synthesis (and transposition) in response to replication of this region during passage of the replicating fork, repair synthesis, or subsequent to conjugal introduction of the sequence.

1.6 CONCLUSION

The transcription initiation process in *E. coli* has been amenable to detailed biochemical studies (largely due to its high fidelity *in vitro*) and genetic analyses. Although there are some interesting areas of uncertainty, these types of studies have allowed us to identify promoters with a high degree of confidence and to obtain an estimate of the activity level of any given promoter. This type of information was used in the design and construction of a totally synthetic promoter (Soberon et al. 1982). We also have information about the steps involved in the transcription initiation process, but we know very little about how these steps are related to the detailed structure(s) of the promoter. Some of the mechanisms that regulate the transcription initiation reaction have been described. The well-known mechanism of negative regulation via a repressor-operator interaction system is the best understood. It can be used effectively to construct regulated expression systems. The overall consequence of positive control systems is also understood, but the mechanisms have not been determined. Other interesting new regulatory schemes have been discovered. In conclusion, we know a great deal about the transcription initiation process, but there is a great deal left to learn.

REFERENCES

Adhya, S., and Miller, W. (1979) *Nature* 279:492–494.

Aiba, H., Adhya, S., and deCrombrugghe, B. (1981) *J. Biol. Chem.* 256:11905–11910.

An, G., and Friesen, J.D. (1979) *J. Bacteriol.* 140:400–407.

Arditti, R.R., Grodzicker, T., and Beckwith, J. (1973) *J. Bacteriol.* 114:652–655.

Barkley, M., and Bourgeois, S. (1978) in *The Operon*, Miller, J., and Reznikoff, W.S., eds. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp. 177–200.

Beckwith, J. (1981) *Cell* 23:307–308.

Beckwith, J., Grodzicker, T., and Arditti, R.R. (1972) *J. Mol. Biol.* 69:155–160.

Beckwith, J., and Zipser, D., eds. (1970) *The Lactose Operon*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Bertrand, K.P., Postle, K., Wray, L.V., Jr., and Reznikoff, W.S. (1983) *Gene* 23:149–156.

Bertrand, K.P., Postle, K., Wray, L.V., Jr., and Reznikoff, W.S. (1984) *J. Bacteriol.* 158:910–919.

Bertrand-Burggraf, E., LeFevre, J.F., and Daune, M. (1984) *Nucleic Acids Res.* 12:1697–1706.

Beyreuther, K. (1978) in *The Operon,* Miller, J., and Reznikoff, W.S., eds. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp. 123–154.
Buc, H., and McClure, W.R. (1984) *Biochemistry* 24:2712–2723.
Burgess, R.R., Travers, A.A., Dunn, J.J., and Bautz, E.K.F. (1969) *Nature* 221:43–46.
Cannistraro, V.J., and Kennell, D.E. (1985) *J. Mol. Biol.* 182:241–248.
Carpousis, A.J.,and Gralla, J.D. (1980) *Biochemistry* 19:3245–3253.
Caruthers, M.H., Beaucage, S.L., Efcavitch, J.W., Fisher, E.F., Goldman, R.A., deHaseth, P.L., Mandecki, W., Matteucci, M.D., Rosendahl, M.S., and Stabinsky, Y. (1982) *Cold Spring Harbor Symp. Quant. Biol.* 47:411–418.
Casadaban, M.J., and Cohen, S.N. (1980) *J. Mol. Biol.* 138:179–207.
Chamberlin, M.J. (1974) *Annu. Rev. Biochem.* 43:721–775.
Cohen, G.N., and Jacob, F. (1959) *C.R. Acad. Sci.* 240:3490.
Crothers, D.M., and Fried, M. (1983) *Cold Spring Harbor Symp. Quant. Biol.* 47:263–269.
deCrombrugghe, B.D., Busby, S., and Buc, H. (1984) in *Biological Regulation and Development*, vol. VI-B, Yamamoto, K., ed. Plenum Press, New York, pp. 129–167.
Dickson, R.C., Abelson, J.N., Barnes, W.M., and Reznikoff, W.S. (1975) *Science* 187:27–35.
Dickson, R.C., Abelson, J.N., Johnson, P., Reznikoff, W.S., and Barnes, W.M. (1977) *J. Mol. Biol.* 111:65–75.
DiLauro, R., Taniguchi, T., Musso, R., and deCrombrugghe, B. (1979) *Nature* 279:494–500.
Emmer, M.B., deCrombrugghe, B., Paston, I., and Perlman, R. (1970) *Proc. Natl. Acad. Sci. USA* 66:480–487.
Englesberg, E., Irr, J., Power, J., and Lee, N. (1965) *J. Bacteriol.* 90:946–957.
Franklin, N.C. (1971) in *The Bacteriophage Lambda*, Hershey, A.D., ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp. 621–638.
Frederick, G.A., Grable, J., Melia, M., Samudzi, C., Jen-Jacobson, L., Wang, B.-C., Greene, P., Boyer, H.W., and Rosenberg, J.M. (1984) *Nature* 309:327–331.
Gamper, H.B., and Hearst, J.E. (1982) *Cell* 29:81–90.
Gellert, M. (1981) *Annu. Rev. Biochem.* 50:879–910.
Gilbert, W. (1976) in *RNA Polymerase*, Losick, R., and Chamberlin, M., eds. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp. 193–206.
Gilbert, W., Gralla, J., Majors, J., and Maxam, A. (1975) in *Protein-Ligand Interactions*, Sund, H., and Blauer, G., eds. Walter de Gruyter, Berlin, pp. 193–210.
Grossman, A., Erickson, J., and Gross, C. (1984) *Cell* 38:383–390.
Guarente, L., Nye, J.S., Hochschild, A., and Ptashne, M. (1982) *Proc. Natl. Acad. Sci. USA* 79:2236–2239.
Hahn, W.E., Pettijohn, D.E., and Van Ness, J. (1977) *Science* 197:582–585.
Hansen, U.M., and McClure, W.R. (1980) *J. Biol. Chem.* 255:9564–9570.
Hawley, D.K., Johnson, A.D., and McClure, W.R. (1985) *J. Biol. Chem.* 260:8618–8626.
Hawley, D.K., and McClure, W.R. (1980) *Proc. Natl. Acad. Sci. USA* 77:6381–6385.
Hawley, D.K., and McClure, W.R. (1982) *J. Mol. Biol.* 157:493–525.
Hawley, D.K., and McClure, W.R. (1983a) *Nucleic Acids Res.* 11:2237–2255.
Hawley, D.K., and McClure, W.R. (1983b) *Cell* 32:327–333.
Hawley, D.K., Malan, T.P., Mulligan, M.E., and McClure, W.R. (1982) in *Promoters: Structure and Function*, Rodriguez, R.L., and Chamberlin, M.J., eds. Praeger, New York, pp. 54–68.
Hillen, W., and Schollmeir, K. (1983) *Nucleic Acids Res.* 11:525–529.
Hochschild, A., Irwin, N.,
Hopkins, J. (1974) *J. Mol. I*
Irani, M.H., Orosz, L., and
Jacob, F., and Monod, J. (1
Kadesch, T.R., Rosenberg,
Kadesch, T.R., Williams, F
Kirkegaard, K., Buc, H., S
USA 80:2544–2548.
Kleckner, N., Morisato, D
Symp. Quant. Biol. 49:
Kolb, A., Spassky, A., Cha
11:7833–7852.
Kunkle, T.A. (1984) *Proc. I*
LeClerk, J.E., and Istock, I
Losick, R., and Pero, J. (19
Maizels, N. (1973) *Proc. Na*
Majors, J. (1975) *Proc. Nat.*
Majors, J. (1977) Ph.D. diss
Malan, T.P., Kolb, A., Buc
Malan, T.P., and McClure,
Mandecki, W., and Caruth
Mandecki, W., and Reznik
Maquat, L.E., and Rezniko
McClure, W.R. (1983) in
Stratman, F.W., and
pp. 207–217.
McClure, W.R. (1985) *Ann*
McClure, W.R., Hawley, I
Function, Rodriguez.
pp. 111–120.
McClure, W.R., Hawley, D
Harbor Symp. Quant.
McKay, D.B., and Steitz, T
McKenney, K., Shimatake,
M. (1982) in *Gene Am*
T.S., eds. Elsevier Nor
Melancon, P., Burgess, R.R
Menzel, R., and Gellert, M
Meyer, B.J., Kleid, D.G.
72:4785–4789.
Miller, J.H. (1978) in *The*
Harbor Laboratory, C
Miller, J.H., and Reznikoff.
tory, Cold Spring Harl
Moran, C.P., Jr., Lang, N.,
Pero, J. and Losick, R
Mulligan, M.E., Brosius, J.
Mulligan, M.E., Hawley, D
Res. 12:789–800.

Hochschild, A., Irwin, N., and Ptashne, M. (1983) *Cell* 32:319–325.
Hopkins, J. (1974) *J. Mol. Biol.* 87:715–724.
Irani, M.H., Orosz, L., and Adhya, S. (1983) *Cell* 32:783–788.
Jacob, F., and Monod, J. (1961) *J. Mol. Biol.* 3:318–356.
Kadesch, T.R., Rosenberg, S., and Chamberlin, M.J. (1982) *J. Mol. Biol.* 155:1–29.
Kadesch, T.R., Williams, R.C., and Chamberlin, M.J. (1980) *J. Mol. Biol.* 136:79–93.
Kirkegaard, K., Buc, H., Spassky, A., and Wang, J.C. (1983) *Proc. Natl. Acad. Sci. USA* 80:2544–2548.
Kleckner, N., Morisato, D., Roberts, D., and Bender, J. (1984) *Cold Spring Harbor Symp. Quant. Biol.* 49:235–244.
Kolb, A., Spassky, A., Chapon, C., Blazy, B., and Buc, H. (1983) *Nucleic Acids Res.* 11:7833–7852.
Kunkle, T.A. (1984) *Proc. Natl. Acad. Sci. USA* 81:1494–1498.
LeClerk, J.E., and Istock, N.L. (1982) *Nature* 297:596–598.
Losick, R., and Pero, J. (1981) *Cell* 25:582–584.
Maizels, N. (1973) *Proc. Natl. Acad. Sci. USA* 70:3585–3589.
Majors, J. (1975) *Proc. Natl. Acad. Sci. USA* 72:4394–4398.
Majors, J. (1977) Ph.D. dissertation, Harvard University.
Malan, T.P., Kolb, A., Buc, H., and McClure, W.R. (1984) *J. Mol. Biol.* 180:881–909.
Malan, T.P., and McClure, W.R. (1984) *Cell* 39:173–180.
Mandecki, W., and Caruthers, M.H. (1984) *Gene* 31:263–267.
Mandecki, W., and Reznikoff, W.S. (1982) *Nucleic Acids Res.* 10:903–912.
Maquat, L.E., and Reznikoff, W.S. (1980) *J. Mol. Biol.* 139:551–556.
McClure, W.R. (1983) in *Biochemistry of Metabolic Processes*, Lennon, D.L.F., Stratman, F.W., and Zahlten, R.N., eds. Elsevier Biomedical, New York, pp. 207–217.
McClure, W.R. (1985) *Ann. Rev. Biochem.* 54:171–204.
McClure, W.R., Hawley, D.K., and Malan, T.P. (1982) in *Promoters: Structure and Function*, Rodriguez, R.L., and Chamberlin, M.J., eds. Praeger, New York, pp. 111–120.
McClure, W.R., Hawley, D.K., Youderian, P., and Susskind, M.M. (1983) *Cold Spring Harbor Symp. Quant. Biol.* 47:477–481.
McKay, D.B., and Steitz, T.A. (1981) *Nature* 290:744–749.
McKenney, K., Shimatake, H., Court, D., Schmeissner, U., Brady, D., and Rosenberg, M. (1982) in *Gene Amplification and Analysis*, vol. 2, Shirkjian, J.G., and Papas, T.S., eds. Elsevier North Holland, New York, pp. 383–415.
Melancon, P., Burgess, R.R., and Record, M.T., Jr. (1982) *Biochemistry* 21:4318–4331.
Menzel, R., and Gellert, M. (1983) *Cell* 34:105–113.
Meyer, B.J., Kleid, D.G., and Ptashne, M. (1975) *Proc. Natl. Acad. Sci. USA* 72:4785–4789.
Miller, J.H. (1978) in *The Operon*, Miller, J., and Reznikoff, W.S., eds. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp. 31–88.
Miller, J.H., and Reznikoff, W.S. eds. (1978) *The Operon*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
Moran, C.P., Jr., Lang, N., LeGrice, S.F.J., Lee, G., Stephens, M., Sonenshein, A.L., Pero, J. and Losick, R. (1982) *Mol. Gen. Genet.* 186:339–346.
Mulligan, M.E., Brosius, J., and McClure, W.R. (1985) *J. Biol. Chem.* 260:3529–3538.
Mulligan, M.E., Hawley, D.K., Entriken, R., and McClure, W.R. (1984) *Nucleic Acids Res.* 12:789–800.

Munson, L.M. (1983) Ph.D. dissertation, University of Wisconsin.
Munson, L.M., and Reznikoff, W.S. (1981) *Biochemistry* 20:2081–2085.
Munson, L.M. Stormo, G.D., Niece, R.L., and Reznikoff, W.S. (1984) *J. Mol. Biol.* 177:663–683.
Musso, R.E., DiLauro, R., Adhya, S., and deCrombrugghe, B. (1977) *Cell* 12:847–854.
Okamoto, T., Sugimoto, K., Sugisaki, H., and Takanami, M. (1977) *Nucleic Acids Res.* 4:2213–2222.
O'Neill, M.C., Amass, K., and deCrombrugghe, B. (1981) *Proc. Natl. Acad. Sci. USA* 78:2213–2217.
Pabo, C.O., and Sauer, R.T. (1984) *Annu. Rev. Biochem.* 53:293–321.
Peterson, M., and Reznikoff, W.S. (1984a) *J. Mol. Biol.* 185:535–543.
Peterson, M., and Reznikoff, W.S. (1984b) *J. Mol. Biol.* 185:525–533.
Prosen, D., and Cech, C.L. (1985) *Biochemistry* 24:2219–2227.
Queen, C., and Rosenberg, M. (1981) *Nucleic Acids Res.* 9:3365–3377.
Raibaud, O., and Schwartz, M. (1984) *Annu. Rev. Genet.* 18:178–206.
Reznikoff, W.S. (1976) in *RNA Polymerase*, Losick, R., and Chamberlin, M., eds. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp. 441–454.
Reznikoff, W.S. (1984) in *The Microbe 1984. II: Prokaryotes and Eukaryotes*, Kelly, D.P., and Carr, N.G., eds. Cambridge University Press, Cambridge, pp. 195–218.
Reznikoff, W.S., Maquat, L.E., Munson, L.M., Johnson, R.C., and Mandecki, W. (1982) in *Promoters: Structure and Function*, Rodriguez, R.L., and Chamberlin, M.J., eds. Praeger, New York, pp. 80–95.
Reznikoff, W.S., Miller, J.H., Scaife, J.G., and Beckwith, J.R. (1969) *J. Mol. Biol.* 43:201–213.
Reznikoff, W.S., Winter, R.B., and Hurley, C.K. (1974) *Proc. Natl. Acad. Sci. USA* 71:2314–2318.
Roe, J.H., Burgess, R.R., and Record, T.R., Jr. (1984) *J. Mol. Biol.* 176:495–521.
Rosenberg, M., and Court, D. (1979) *Annu. Rev. Genet.* 13:319–353.
Sanzey, B. (1979) *J. Bacteriol.* 138:40–47.
Scaife, J., and Beckwith, J. (1967) *Cold Spring Harbor Symp. Quant. Biol.* 31:403–408.
Schmeissner, V., Court, D., Shimatake, H., and Rosenberg, M. (1980) *Proc. Natl. Acad. Sci. USA* 77:3191–3195.
Schwartz, D., and Beckwith, J. (1970) in *The Lactose Operon*, Beckwith, J., and Zipser, D., eds. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp. 417–422.
Shaner, S.L., Melancon, P., Lee, K.S., Burgess, R.R., and Record, M.T., Jr. (1983) *Cold Spring Harbor Symp. Quant. Biol.* 47:463–472.
Shih, M.-C., and Gussin, G.N. (1983a) *Proc. Natl. Acad. Sci. USA* 80:496–500.
Shih, M.-C., and Gussin, G.N. (1983b) *Cell* 34:941–949.
Shih, M.-C., and Gussin, G.N. (1984) *J. Mol. Biol.* 172:489–506.
Siebenlist, U. (1979) *Nature* 279:651–652.
Siebenlist, U., Simpson, R.B., and Gilbert, W. (1980) *Cell* 20:269–281.
Silverstone, A., Arditti, R.R., and Magasanik, B. (1970) *Proc. Natl. Acad. Sci. USA* 66:773–779.
Simons, R.W., Hoopes, B.C., McClure, W.R., and Kleckner, N. (1983) *Cell* 34:673–682.
Soberon, X., Rossi, J.J., Larson, G.P., and Itakura, K. (1982) in *Promoters: Structure and Function*, Rodriguez, R., and Chamberlin, M., eds. Praeger, New York, pp. 407–431.
Spassky, A., Busby, S., and Buc, H. (1984) *EMBO J.* 3:43–50.
Squires, C., Lee, F., Bert
(1976) *J. Mol. Biol.* 10
Staden, R. (1984) *Nucleic*
Steege, D.A. (1977) *Proc.*
Stefano, J.E., and Gralla,
Stefano, J.E., and Gralla,
Taniguchi, T., O'Neill, M.,
76:5090–5094.
Travers, A.A., Lamond, A.
vonHippel, P.H., Bear, D.
Biochem. 53:389–446.
Walter, G., Zillig, W., Paln
Ward, D.F., and Murray, N
Weber, K., and Geisler, N.
Cold Spring Harbor L
Wertman, K.F., Little, J.
81:3801–3805.
West, R.W., Neve, R.L., a
Wood, D.C., and Lebowitz
Wu, H.-M., and Crothers,
Yin, J.C.-P., and Reznikoff
Yu, X.-M., Munson, L., an
Yu, X.-M., and Reznikoff,
Zieg, J., Silverman, M., H

Squires, C., Lee, F., Bertrand, K., Squires, C.L., Bronson, M.J., and Yanofsky, C. (1976) *J. Mol. Biol.* 103:351–389.
Staden, R. (1984) *Nucleic Acids Res.* 12:505–519.
Steege, D.A. (1977) *Proc. Natl. Acad. Sci. USA* 74:4163–4167.
Stefano, J.E., and Gralla, J.D. (1982a) *J. Biol. Chem.* 257:13924–13929.
Stefano, J.E., and Gralla, J.D. (1982b) *Proc. Natl. Acad. Sci. USA* 70:1069–1072.
Taniguchi, T., O'Neill, M., and deCrombrugghe, B. (1979) *Proc. Natl. Acad. Sci. USA* 76:5090–5094.
Travers, A.A., Lamond, A.I., Mace, H.A.F., and Berman, M.L. (1983) *Cell* 35:265–273.
vonHippel, P.H., Bear, D.G., Morgan, W.D., and McSwiggen, J.A. (1984) *Annu. Rev. Biochem.* 53:389–446.
Walter, G., Zillig, W., Palm, P., and Fuchs, E. (1967) *Eur. J. Biochem.* 3:194–201.
Ward, D.F., and Murray, N.E. (1979) *J. Mol. Biol.* 133:249–266.
Weber, K., and Geisler, N. (1978) in *The Operon*, Miller, J., and Reznikoff, W.S., eds. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp. 155–176.
Wertman, K.F., Little, J.W., and Mount, D.W. (1984) *Proc. Natl. Acad. Sci. USA* 81:3801–3805.
West, R.W., Neve, R.L., and Rodriguez, R.L. (1979) *Gene* 7:271–288.
Wood, D.C., and Lebowitz, J. (1984) *J. Biol. Chem.* 259:11184–11187.
Wu, H.-M., and Crothers, D. (1984) *Nature* 308:509–513.
Yin, J.C.-P., and Reznikoff, W.S. (1986), in preparation.
Yu, X.-M., Munson, L., and Reznikoff, W.S. (1984) *J. Mol. Biol.* 172:355–362.
Yu, X.-M., and Reznikoff, W.S. (1984) *Nucleic Acids Res.* 12:5449–5464.
Zieg, J., Silverman, M., Hilmen, M., and Simon, M. (1977) *Science* 196:170–172.

Cell, Vol. 62, 1227–1233, September 21, 1990, Copyright © 1990 by Cell Press

Correction of the Cystic Fibrosis Defect In Vitro by Retrovirus-Mediated Gene Transfer

Mitchell L. Drumm,*† Heidi A. Pope,*‡
William H. Cliff,§ Johanna M. Rommens,‖
Sheila A. Marvin,§ Lap-Chee Tsui,‖
Francis S. Collins,*† Raymond A. Frizzell,§ and
James M. Wilson*‡
*Department of Internal Medicine
†Department of Human Genetics
‡Department of Biological Chemistry
Howard Hughes Medical Institute
University of Michigan Medical School
Ann Arbor, Michigan 48109-0650
§Gregory Fleming James Cystic Fibrosis Research Center
and Department of Physiology and Biophysics
University of Alabama at Birmingham
UAB Station
Birmingham, Alabama 35294
‖Department of Genetics
Research Institute
The Hospital for Sick Children
Toronto, Ontario M5G 1X8
Canada

Summary

We have used retrovirus-mediated gene transfer to demonstrate complementation of the cystic fibrosis (CF) defect in vitro. Amphotropic retroviruses were used to transduce a functional cystic fibrosis transmembrane conductance regulator (CFTR) cDNA into CFPAC-1, a pancreatic adenocarcinoma cell line derived from a patient with CF that stably expresses the chloride transport abnormalities characteristic of CF. CFPAC-1 cells were exposed to control virus (PLJ) and CFTR-expressing virus (PLJ-CFTR); viral-transduced clones were isolated and subjected to molecular and physiologic analysis. RNA analysis detected a viral-derived CFTR transcript in all of the PLJ-CFTR clones that contained unrearranged proviral sequences. Agents that increase intracellular cAMP stimulated ^{125}I efflux in PLJ-CFTR clones but not PLJ clones. Whole-cell patch-clamp performed on three responding clones showed that the anion efflux responses were due to cAMP stimulation of Cl conductance. Our findings indicate that expression of the normal CFTR gene confers cAMP-dependent Cl channel regulation on CF epithelial cells.

Introduction

Cystic fibrosis (CF) is an autosomal recessive disorder characterized by abnormalities in water and electrolyte transport that lead to pancreatic and pulmonary insufficiency (Taussig, 1984). Recent progress in our understanding of the genetic and functional basis of CF has provided a foundation for better defining its molecular pathology as well as developing novel therapies based on somatic gene transfer.

Functional expression of the CF defect reduces the chloride ion permeability of epithelial tissues (Quinton, 1990). The ability of epithelial cells in the airways, sweat glands, pancreas, and other tissues to secrete Cl in response to cAMP-mediated agonists is lost or severely reduced. Activation of apical membrane Cl channels by cAMP-dependent protein kinase is impaired, but channels with normal conductance properties can be activated by other means, including agonists whose effects are mediated by increased cell Ca (Frizzell, 1987; Welsh, 1990). These findings have led to the suggestion that the Cl channel per se is not defective in CF, but that the defect may lie in a regulatory protein that transduces the effects of protein kinase activation. Abnormalities in epithelial sodium transport in CF cells (Boucher et al., 1986) support the concept of a regulatory defect that can affect other cellular functions.

Isolation of the gene for CF (Rommens et al., 1989; Riordan et al., 1989; Kerem et al., 1989) has provided further insight into the molecular basis of the disease. The gene responsible for CF was localized to 250,000 bp of genomic DNA based on its location within the genome. It encodes a protein of 1,480 amino acids called the cystic fibrosis transmembrane conductance regulator (CFTR) (Riordan et al., 1989). The most compelling evidence thus far to support the role of CFTR in the etiology of CF has been provided by genetic analyses (Kerem et al., 1989). Sequence analysis of the CFTR gene of CF chromosomes has revealed a variety of mutations (Cutting et al., 1990; White et al., 1990; Dean et al., 1990; Kerem et al., 1990). Extensive population studies have indicated that the most common CF mutation, a deletion of the 3 nucleotides that encode phenylalanine 508 (ΔF_{508}), is present on 70% of all CF chromosomes, but never on normal chromosomes (Kerem et al., 1989; Cystic Fibrosis Genetic Analysis Consortium, 1990).

The specific role that CFTR plays in Cl transport remains unclear. The CFTR protein contains several interesting functional domains including two nucleotide binding folds, a regulatory region that has many possible sites for phosphorylation, and two hydrophobic regions that probably interact with cell membranes. CFTR shows structural similarity with several members of the "ATP binding cassette" (ABC) superfamily of proteins, including the periplasmic binding proteins of prokaryotes and the P glycoprotein associated with multidrug resistance in higher eukaryotes (Riordan et al., 1989; Hyde et al., 1990).

As the next step in understanding the role of CFTR in the dysregulation of Cl channel activation in CF, we have achieved complementation of the CF defect in vitro by retrovirus-mediated transduction of a full-length CFTR cDNA.

Results

Recombinant Retroviruses

Early attempts to reconstitute a full-length CFTR cDNA from overlapping clones were uniformly unsuccessful.

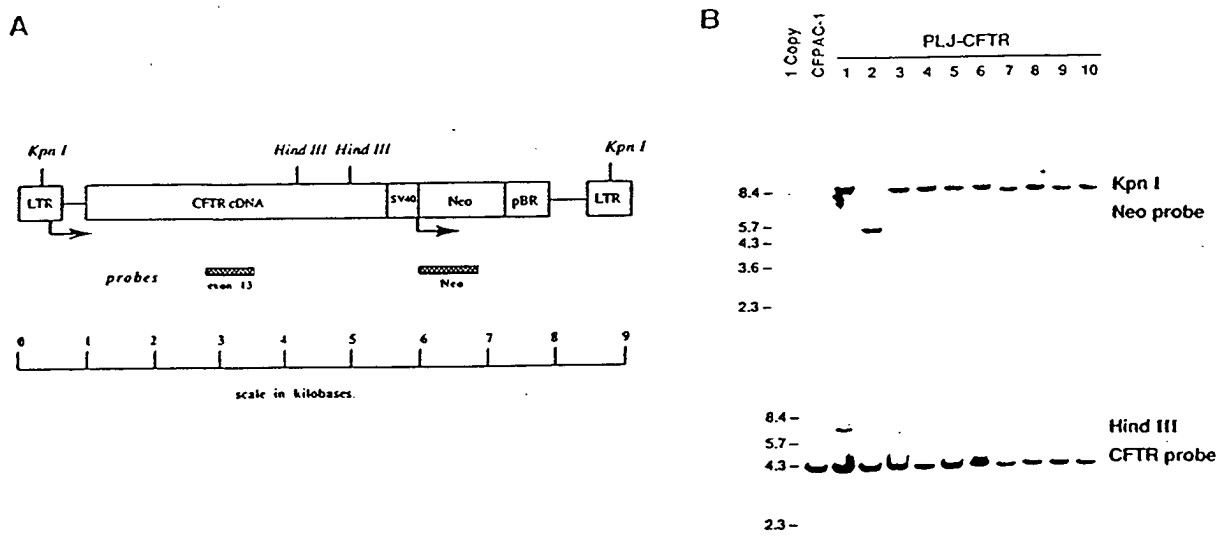


Figure 1. Analysis of Proviral Sequences in Transduced CFPAC-1 Clones

(A) Retroviral vector PLJ-CFTR. Important structural components of the vector include the LTR sequences, CFTR cDNA, sequences from the origin of SV40, the gene that confers resistance to G418 (Neo), and the origin of replication for pBR322 (pBR). Sites of transcriptional initiation are indicated with arrows at the 5′ LTR and internal SV40 sequences. Recognition sites for restriction endonucleases KpnI and HindIII are indicated. Probes specific for the CFTR gene (exon 13) and the Neo gene that were used in DNA and RNA blot analysis are indicated below the vector.

(B) CFPAC-1 cells were infected with PLJ or PLJ-CFTR virus and selected in the presence of G418 in order to isolate individual clones. High molecular weight DNA was isolated from each clone and analyzed by gel blot hybridization. In the top panel, DNA was digested with KpnI and the filter was hybridized to a Neo-specific probe; in the bottom panel, DNA was digested with HindIII and the filter was hybridized to the exon 13 CFTR-specific probe. The 4.3 kb band in all lanes arises from the endogenous CFTR gene. Samples include: lane CFPAC-1, CFPAC-1 DNA (10 μg); lane 1 Copy, CFPAC-1 DNA (10 μg) supplemented with 7.5 pg of PLJ-CFTR plasmid DNA; and lanes numbered 1–10, DNA (10 μg) from PLJ-CFTR clones 1 through 10. The viral band comigrates with the endogenous band in clone 5; a unique integrant was clearly demonstrated in this clone when the filter was stripped and rehybridized with a Neo probe. Molecular size standards in kilobases are indicated at left.

The exact cause of these difficulties remains to be defined, but we have data to show that prokaryotic transcription from internal CFTR cDNA sequences may result in the expression of a protein that is toxic to bacteria. The introduction of three silent mutations (T to C at 930, A to G at 933, and T to C at 936) into a restriction fragment of CFTR that spans exon 6b completely ablated this toxic effect, potentially by interfering with the cryptic prokaryotic promoter, and enabled the reconstruction of 4.6 kb of contiguous CFTR cDNA sequence. The nucleotide sequence of this reconstructed cDNA was redetermined and found identical to that published previously (Riordan et al., 1989), with the exception of the three silent mutations noted above.

The modified CFTR cDNA was cloned into the previously described retroviral vector PLJ (Korman et al., 1987). The proviral component of this recombinant vector, called PLJ-CFTR, is depicted in Figure 1A. Transcription from the 5′ long terminal repeat (LTR) produces an 8.5 kb genomic transcript that is responsible for viral passage and CFTR expression. Transcription from SV40 sequences located internal to the LTRs leads to the formation of a second transcriptional unit that expresses a Neo selectable marker.

Transfection of PLJ and PLJ-CFTR vectors into the virus packaging cell line ψCrip led to the transient production of replication-defective virus. Limiting dilutions of virus stocks were used to infect CFPAC-1 cells, which were subsequently cultured in the presence of G418 in order to select for transduced clones. Transiently produced PLJ-CFTR virus stocks had a lower titer (50 to 100-fold) than those produced with the PLJ vector. Ten individual clones of cells were isolated from infections performed with each type of virus (named PLJ clones 1 through 10 and PLJ-CFTR clones 1 through 10) and subjected to molecular and physiologic analyses.

Transduced Clones Express Retroviral CFTR Sequences

Retrovirally transduced clones of CFPAC-1 cells were analyzed for proviral sequences as described for other cell types (Wilson et al., 1988, 1990a, 1990b). Digestion of high molecular weight DNA with a restriction enzyme, KpnI, that has unique sites in the vector LTRs releases all integrated forms of the PLJ-CFTR provirus as a common 8.5 kb fragment. Gel blot hybridization of KpnI-restricted DNA revealed unrearranged proviral sequences with the expected abundance of 1 copy per cell in 10/10 PLJ clones and 9/10 PLJ-CFTR clones (Figure 1B, top panel). Hybridization of the filter with a Neo-specific probe detected a markedly rearranged provirus in PLJ-CFTR clone 2; this virus apparently deleted a major part of the CFTR cDNA (data not shown). Gel blot hybridization analysis was also used to study the complexity and uniqueness of each

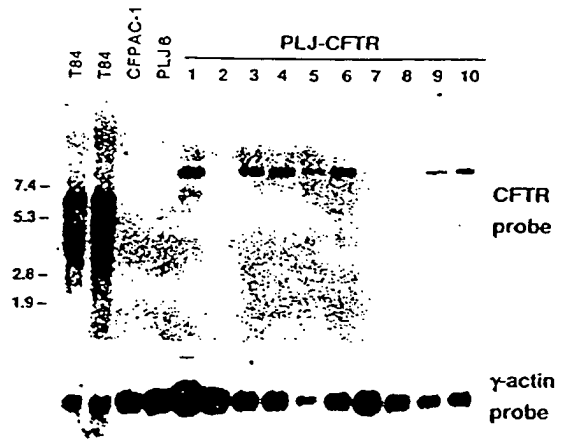


Figure 2. RNA Blot Analysis of CFPAC-1 Clones

Clones of retrovirus-transduced CFPAC-1 cells were isolated and analyzed for the presence of CFTR transcripts. Total cellular RNA was harvested from individual clones and subjected to RNA blot analysis using the exon 13 CFTR probe to hybridize with the filter (top panel). The filter was stripped and rehybridized with a probe derived from human γ-actin cDNA (bottom panel) to control for variation in sample loading. RNA samples (10 μg) were derived from the following cells: lanes T84, duplicate samples from the colonic tumor cell line T84; lane CFPAC-1, nontransduced CFPAC-1 cells; lane PLJ 6, CFPAC-1 clone 6 from the PLJ infection; and lanes numbered 1–10, CFPAC-1 clones 1 through 10 from the PLJ-CFTR infection. Molecular size markers in kilobases are noted at left.

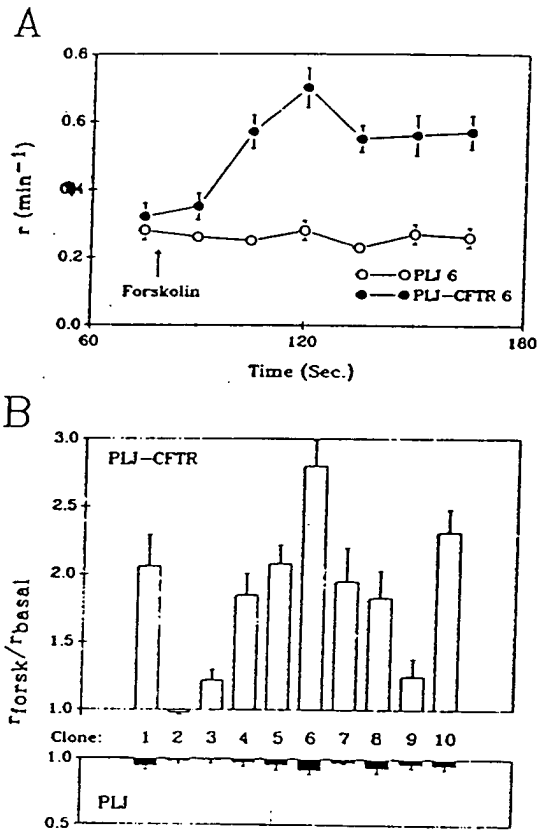


Figure 3. Effect of Forskolin on ^{125}I Efflux in PLJ and PLJ-CFTR Clones

(A) Time course of the efflux rate constant in clones 6. Forskolin (10 μM) was added at the indicated time. The first 60 s of efflux allows for washout of extracellular ^{125}I and is not shown (see Experimental Procedures). Values are means ± SEM of n = 9 determinations for each clone.

(B) Ratio of forskolin-stimulated to basal ^{125}I efflux in PLJ and PLJ-CFTR clones 1–10. r values were taken before and after forskolin addition. For PLJ-CFTR clone 2, the same scaling applies below 1.0. Values are means ± SEM of n = 9 determinations for all clones except PLJ 5, where n = 7.

putative PLJ-CFTR clone. High molecular weight DNA was digested with HindIII, a restriction enzyme with two internal sites in PLJ-CFTR, and analyzed with the exon 13 CFTR-specific probe (Figure 1B, bottom panel). This analysis demonstrated the existence of a single, unique integration site in 9/10 PLJ-CFTR clones; the CFTR-specific probe failed to detect the provirus in DNA from PLJ-CFTR clone 2 because of the apparent deletion described above.

Expression of the retrovirally transduced CFTR gene was studied by RNA blot analysis using the CFTR exon 13 probe (Figure 2). Total cellular RNA from the previously described human colon tumor cell line, T84, demonstrated high levels of the endogenous CFTR transcript. No CFTR transcript was detected by Northern analysis in mock-infected CFPAC-1 cells or PLJ clone 6, though CFTR RNA can be detected in CFPAC-1 by RNA-PCR. A viral-directed CFTR transcript of the expected size (i.e., 8.5 kb) was detected in 9/10 PLJ-CFTR clones; the CFTR probe failed to detect a transcript in RNA from the clone that contains the deleted provirus (PLJ-CFTR clone 2).

Transduced Clones Show Forskolin Stimulation of Anion Transport

Isotopic anion (^{125}I) effluxes were measured to screen the PLJ and PLJ-CFTR clones for cAMP- and Ca-stimulated anion transport. The efflux assay provides a qualitative estimate of agonist-stimulated Cl conductance pathways in Cl-secreting epithelia, as judged from the inhibitory effects of Cl channel blockers and depolarizing membrane potentials on ^{125}I efflux (Venglarik et al., 1990). Figure 3A shows the time course of the ^{125}I efflux rate constant (r) in two clones, PLJ 6 and PLJ-CFTR 6, with and without the addition of forskolin, an agent that stimulates adenylate cyclase. Following a basal efflux period in the absence of agonist, forskolin increased the ^{125}I efflux rate from PLJ-CFTR clone 6 from 0.32 to 0.70 min^{-1}; PLJ 6 did not respond. r values obtained before forskolin addition and during the peak of the forskolin response provided an estimate of the relative stimulation of ^{125}I efflux (i.e., r_{forsk}/r_{basal}). In the responding PLJ-CFTR clones, the peak forskolin effect on anion efflux was observed during the first three flux periods following forskolin addition (15–45 s).

Data derived from 20 clones are illustrated in Figure 3B. Seven of ten PLJ-CFTR clones showed significant increases in ^{125}I efflux in response to forskolin, whereas none (0/10) of the control PLJ clones responded to forskolin. The parent cell line, CFPAC-1, also shows no response to forskolin or cAMP analogs (Schoumacher et al., 1990). PLJ-CFTR clone 2 showed a major deletion in its CFTR cDNA by gel

blot hybridization (Figure 1B), accounting for the failure of forskolin to stimulate ^{125}I efflux. In the seven responding PLJ-CFTR clones, the relative stimulation of anion efflux by forskolin ranged from 1.8 to 2.8-fold. This compares well with the 3.5-fold stimulation of efflux reported recently for the colonic tumor cell line T84 (Venglarik et al., 1990) and with the 2.9-fold stimulation by forskolin observed in primary cultures of normal human trachea (Li et al., 1988). The time course of stimulation in the PLJ-CFTR 6 clone is also similar to that observed in T84. Our results indicate that expression of CFTR cDNA endows CFPAC-1 cells with cAMP-responsive anion efflux.

The correlation between forskolin responsiveness of the PLJ-CFTR clones and their CFTR mRNA levels was not striking (compare Figures 2 and 3B). Three of the best responders in the efflux assay showed high mRNA levels (i.e., PLJ-CFTR clones 1, 6, and 10). In other instances, however, the correlation was not as good. For example, clones 7 and 8 showed approximately a 2-fold response to forskolin but had relatively low mRNA levels, and clones 3 and 9 showed a low forskolin response despite the presence of readily detectable CFTR mRNA.

Addition of the Ca ionophore ionomycin increased ^{125}I efflux in all control and CFTR clones. Values of r_{iono}/r_{basal} averaged 14 ± 2 in PLJ and 14 ± 1 in PLJ-CFTR (n = 20) in each group; no significant differences were detected between individual clones. The extent of response of PLJ clones to ionomycin is similar to that observed previously in wild-type CFPAC-1 cells (Schoumacher et al., 1990) and is about 3 times the response of T84 cells (Venglarik et al., 1990). The ability of Ca ionophores and Ca-mediated agonists to stimulate Cl secretion has been reported for airway and sweat gland cells derived from both normal individuals and CF patients (Sato and Sato, 1984; Frizzell et al., 1986; Willumsen and Boucher, 1989). The presence of this response in CF cells indicates that CFTR is not required for Ca-mediated stimulation of Cl transport. The lack of significant differences in the extent of Ca stimulation in PLJ and PLJ-CFTR clones suggests that CFTR does not modulate the activity of Ca-mediated regulatory pathways that govern Cl secretion.

Clones Transduced with the CFTR Retrovirus Show cAMP-Induced Cl Currents

Whole-cell patch-clamp recordings were used to determine whether the cAMP-induced increase in anion efflux in PLJ-CFTR clones (Figure 3) was due to stimulation of Cl conductance pathways (Cliff and Frizzell, 1990). A typical response of PLJ-CFTR clone 1 is illustrated in Figure 4A. Chloride currents were measured as the inward current produced by voltage pulses to −84 mV. Similar increases in inward current were observed in 11 of 13 cells from PLJ-CFTR clones 1, 6, and 10, in which addition of forskolin (5 μM) or cAMP (200–800 μM) increased inward currents from 220 ± 68 pA to 1690 ± 495 pA in responding cells. The magnitude of this response compares favorably with that observed in T84 cells (Cliff and Frizzell, 1990) and primary tracheal cell cultures derived from normal subjects (W. H. C. and R. A. F., unpublished data). As shown in Figure 4B, no current responses were observed

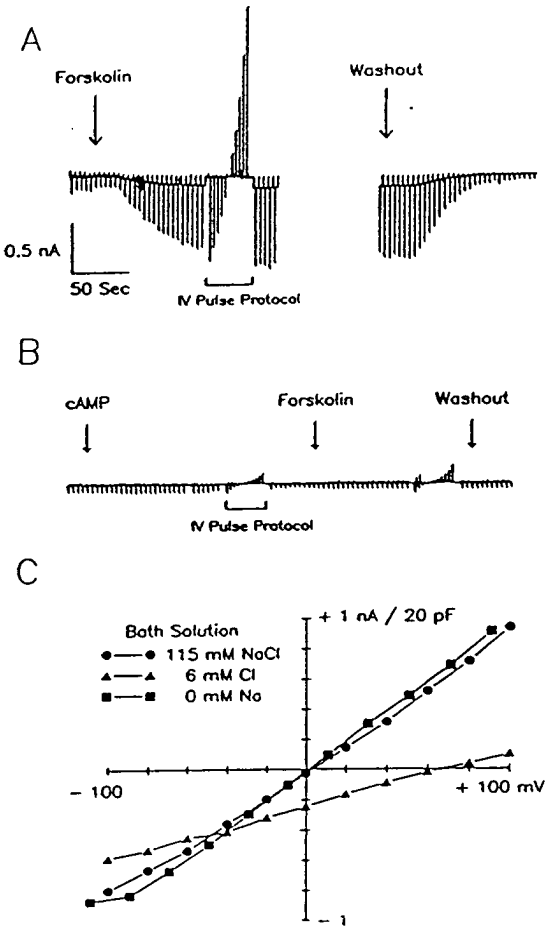


Figure 4. Whole-Cell Cl Currents in PLJ and PLJ-CFTR Clones

(A) Stimulation of inward currents in a PLJ-CFTR clone 1 cell by 5 μM forskolin under whole-cell voltage-clamp. Membrane voltage was held at −10 mV and pulsed to 0 and −84 mV. The gap in the record represents time (6 min) during which bath solution substitutions were performed to determine ion selectivity of the forskolin-induced current (see [C]). Pulse protocols for determining the I–V relations were run at the indicated times. Similar results were obtained in 11 PLJ-CFTR clone 1, 6, and 10 cells.

(B) Failure of cAMP (400 μM) or forskolin (5 μM) to stimulate membrane currents in a PLJ clone 6 cell. Membrane voltage was held at −20 mV and pulsed to 0 mV and −84 mV. Similar results were obtained in six PLJ clone 6 cells.

(C) Instantaneous I–V relations of forskolin-induced currents in NaCl bath, low Cl bath, and Na-free bath. Forskolin-induced currents were obtained by digital subtraction of currents before and after stimulation; the values shown were recorded 6 ms after the initiation of voltage pulses. These data were obtained from the PLJ-CFTR clone 6 cell record shown in (A) during the 6 min recording gap.

in cells from the control clone, PLJ 6 (n = 6). As observed for the ^{125}I efflux determinations, ionomycin (2 μM) increased inward currents in both PLJ (n = 4) and PLJ-CFTR (n = 3) clones.

Figure 4C illustrates current–voltage (I–V) relations of the forskolin-stimulated current obtained from PLJ-CFTR clone 1. The I–V relation of the stimulated current appeared to be linear, as observed in T84 cells (Cliff and Frizzell, 1990). Currents determined using equal bath and pipette Cl concentrations reversed near the Cl equilibrium

potential of 0 mV. Reducing the bath Cl concentration to 6 mM (glutamate replacement) decreased the outward currents and shifted the reversal potential for current flow to +66 mV, a value close to the Cl equilibrium potential (+80 mV) for this outwardly directed Cl gradient. Replacement of bath Na by N-methyl-D-glucamine (NMDG) did not significantly alter the I–V relation. These findings indicate that the forskolin-stimulated current is Cl selective and that the stimulation of an anion efflux in PLJ-CFTR clones is due to activation of Cl conductance pathways.

Discussion

Our findings indicate that expression of the normal CFTR gene from a retroviral vector confers cAMP-dependent Cl channel regulation on CF epithelial cells. cAMP and forskolin increased anion efflux and Cl currents in CFTR-transduced CFPAC-1 clones but not in control clones. The time course and magnitude of stimulation in responding PLJ-CFTR clones are similar to those observed previously in normal human tracheal cells and T84 cells, which express high levels of wild-type CFTR mRNA relative to other cell lines and tissues (Riordan et al., 1989). This suggests that CF cells are capable of transport responses that are physiologically significant in magnitude, provided that the normal gene can be expressed at reasonably high levels.

Evidence that CFTR is in fact the gene for CF was, until now, based solely on genetic data. Numerous CFTR mutations have been identified that are found only on CF chromosomes. Most notable is the common ΔF_{508} deletion, which has now been found on more than 9000 CF chromosomes (Cystic Fibrosis Genetic Analysis Consortium, 1990). Haplotype analyses suggest that the chromosomes carrying ΔF_{508} may have descended from a common ancestor (Kerem et al., 1989). It could be formally argued that this particular mutation was present on the ancestral CF chromosome and is a private polymorphism closely linked to the true CF mutation. This hypothesis would seem unlikely, however, based on the absence of ΔF_{508} on other normal chromosomes of the same haplotype, and the large number of additional mutations (including nonsense and frameshift abnormalities) subsequently identified in CF chromosomes (Cutting et al., 1990; White et al., 1990; Kerem et al., 1990; Dean et al., 1990). The functional data presented here formally prove that CFTR is the CF gene; the addition of a single normal copy of CFTR to a CF cell alleviates the defect in Cl channel regulation, as would be expected for a recessive trait such as CF. Furthermore, the complementation scheme described here provides an assay to determine the validity of other putative CF mutations and to study the function of CFTR by site-specific mutagenesis or domain switching with other members of this large ABC gene family (Hyde et al., 1990).

Preliminary evidence has suggested the possibility of alternative splicing of several exons in CFTR mRNA (T. Strong and F. S. C., unpublished data). Alternative transcriptional start sites have also been detected that could result in different amino termini (J. Koh and F. S. C., unpublished data; J. M. R. and L.-C. T., unpublished data). The possibility that a combination of several CFTR protein isoforms might be necessary for normal Cl channel regulation has thus been raised. The success of the experiments described in this report demonstrates that the cDNA utilized here has all the sequences necessary to encode a functional protein, at least as assayed in CFPAC-1 cells by the analysis of cAMP-stimulated Cl current. Whether other forms of CFTR are also capable of complementation, and whether all aspects of the CF phenotype can be complemented by the cDNA used here, remains to be determined.

Results from both physiologic and molecular cloning studies have raised the possibility that CFTR is a Cl channel. The defect in Cl channel activation by cAMP-dependent protein kinase is present at the single-channel level in cell-free membrane patches (Schoumacher et al., 1987; Li et al., 1988), and the protein structure predicted from CF gene cloning suggests that CFTR is an integral membrane protein with 12 membrane-spanning domains (Riordan et al., 1989). However, in our study there is not a good correlation between the cAMP responsiveness of PLJ-CFTR clones and their CFTR mRNA levels. If the extent of CFTR expression corresponds to its mRNA levels, the lack of correlation between expression and transport suggests that a cellular component(s) other than CFTR is required for cAMP-stimulated anion transport. That is, CFTR is necessary but not sufficient for the Cl channel response, suggesting that CFTR is probably not a cAMP-dependent protein kinase–activatable Cl channel. We cannot exclude the possibility that clonal variabilities in other limiting factors account for this lack of correlation.

CF-associated alterations in other cellular processes such as amiloride-sensitive Na transport (Boucher et al., 1986) and mucin sulfation (Cheng et al., 1989) have been identified. This has given rise to the concept that CFTR may regulate several cellular processes. The role that CFTR plays in Cl channel regulation and in controlling other cellular events has not been elucidated, but the clonal PLJ cell lines will be useful for detailed studies of CFTR function.

Our results also suggest a strategy for treating CF based on retrovirus-mediated transfer of a normal CFTR gene into CF epithelial cells. We have shown that transduction of a single copy of the recombinant CFTR gene leads to stable correction of the Cl channel defect in pancreatic cells and that retroviral vectors are a useful means of transfer despite the large size of the CFTR cDNA. Airway epithelial cells, however, would be the most desirable targets for gene transfer because the pulmonary complications of CF are usually its most morbid and life limiting (Taussig, 1984). Airway epithelial cells are easily infected with recombinant retroviruses (M. C. Iannuzzi, M. L. D., and J. M. W., unpublished data), suggesting that the approach described in this report may also be useful for gene therapies directed to the lung.

Despite the encouraging data on functional complementation of CF presented in this study, many major questions remain about the feasibility of gene therapy of CF using a retrovirus. For example, what type of airway cell

is the appropriate target for gene transfer, and what are the consequences of ectopic expression or overexpression of CFTR? Is it possible to achieve a therapeutic effect if the efficiency of gene transfer is less than 100%? What is the best strategy for obtaining stable and efficient reconstitution of function in the airway in vivo without exposing the patient to excessive risks? The answers to these and other pressing questions in CF biology will need to be intensively sought in order to translate this retroviral complementation strategy into clinical practice.

Experimental Procedures

Cell Lines

CFPAC-1 cells were maintained in culture as described previously (Schoumacher et al., 1990); cells used for retroviral infection were at passage 72. Infected populations of CFPAC-1 cells were selected in medium containing G418 (1 mg/ml) in order to isolate individual clones. Transduced CFPAC-1 cells were removed from selection soon after they were expanded as clones. This was not associated with an apparent loss of proviral sequences or proviral expression. The amphotropic packaging cell line ψCrip was maintained in Dulbecco's modified Eagle's medium supplemented with 10% calf serum and penicillin–streptomycin as described (Danos and Mulligan, 1988).

Construction of CFTR cDNA

The cDNA was constructed by joining the overlapping clones 10-1, T16-1, and T16-4.5 (Riordan et al., 1989). Clones 10-1 and T16-1 were ligated at the unique NruI site in exon 4, and the resultant construct, spanning exons 1 through 13, was joined to T16-4.5. This was done by inserting a SacI–EcoRI partial digestion product of T16-4.5, extending from exon 13 to exon 24, into the respective sites of the 5′ 13-exon construct. These manipulations generated a 4.5 kb clone containing the entire coding sequence as described (Riordan et al., 1989). It was observed that most clones generated from these construction attempts were grossly rearranged. Upon sequencing of an apparently intact construct, a 57 bp deletion was identified in exon 6b occurring between the two copies of a 13 bp direct repeat. On inspection, this interval was noted to contain a consensus prokaryotic promoter sequence. In an attempt to disrupt the repeat, three single-nucleotide alterations were made by in-vitro mutagenesis (J. M. R. and L.-C. T., unpublished data). The introduced changes, which do not alter the CFTR translation product but result in a stable construct, include substitution of T for C at position 930, A for G at 933, and T for C at position 936. The modified reconstructed CFTR plasmid is called CFTR4.6

Retroviral Vectors and Recombinant Retroviruses

Digestion of the modified CFTR plasmid with SacI released the modified CFTR cDNA on a 4.6 kb restriction fragment. The SacI sites were converted to BclI sites with oligonucleotides, and the linkered fragment was cloned into the BamHI site of the previously described retroviral vector PLJ (Korman et al., 1987). This recombinant vector, called PLJ-CFTR, is presented in Figure 1A. Retroviral vectors PLJ and PLJ-CFTR were transfected into the amphotropic packaging cell line ψCrip as described. Tissue culture medium was removed from plates containing the transfected packaging cells 24 hr later in order to harvest the transiently produced amphotropic virus.

CFPAC-1 cells, passaged 1:5 onto 10 cm^2 plates, were exposed to viral supernatants supplemented with Polybrene (4 μg/ml) for 12 to 16 hr. When the cells reached confluence, they were passaged 1:10 into medium containing G418 (1 mg/ml). Clones of cells were isolated, expanded, and cryopreserved.

DNA and RNA Analysis of CFPAC-1 Clones

High molecular weight DNA was isolated from CFPAC-1 cells as described and analyzed by gel blot hybridization (Wilson et al., 1988). Total cellular RNA was purified and subjected to RNA blot analysis (Wilson et al., 1988). Filters were hybridized with a variety of DNA probes that were labeled to a high specific activity using the random priming method (Feinberg and Vogelstein, 1983). These probes include the following: exon 13 of CFTR, isolated following PCR amplification of cloned cDNA using oligonucleotides that flank the borders of this exon (nucleotides 1900 to 2611); Neo-specific sequences on a 960 bp HindIII–NcoI fragment of pSV2Neo; and human γ-actin cDNA.

Anion Efflux Measurements

Radioisotopic anion efflux was determined as described (Venglarik et al., 1990). Briefly, cell monolayers were preloaded with ^{125}I for 30 min; after two washes, efflux was monitored at 15 s intervals using a sample–replace procedure. At the end of the experiment, tracer remaining in the cell monolayer was extracted with 0.1 N NPO_3. The efflux rate constant (r) for each sampling interval was calculated as follows: $r = [\ln(R_1) - \ln(R_2)]/(t_1 - t_2)$, where R_1 and R_2 are the percent of loaded ^{125}I remaining in the monolayer at times t_1 and t_2. Forskolin or ionomycin was added after the fifth 15 s sampling interval. The degree of agonist stimulation is expressed as $r_{agonist}/r_{basal}$, where $r_{agonist}$ is the maximal value observed in the presence of agonist and r_{basal} is taken from the flux interval immediately prior to agonist addition.

Most of the extracellular ^{125}I washout occurs during the initial 60 s of sampling (Venglarik et al., 1990); this period was ignored in the rate constant calculations. However, a small residual efflux from the extracellular space after 60 s leads to a slight underestimate of the agonist response because the extracellular compartment washes out faster than the cellular compartment. Therefore, when there is no efflux response to forskolin, r determined immediately after forskolin addition is slightly less than that measured before forskolin is added. This accounts for the finding that r_{forsk}/r_{basal} is between 0.9 and 1.0 in the PLJ clones (Figure 3B).

Whole-Cell Current Recordings

Macroscopic currents were recorded during whole-cell patch-clamp by methods previously described (Cliff and Frizzell, 1990). Recordings were made at 37°C with the following solutions (mM): bath 115 NaCl, 40 NMDG–glutamate, 5 potassium glutamate, 2 $MgCl_2$, 1 $CaCl_2$, 10 HEPES (pH 7.2); pipette 115 KCl, 35 NMDG–glutamate, 0.25 EGTA, 0.09 $CaCl_2$ (100 nM free Ca), 2 $MgCl_2$, 2 Na_2ATP, 0.20 Na_2GTP, 10 HEPES (pH 7.2). Membrane potentials were clamped alternately for 500 ms durations at three voltages, two of which were chosen to equal the equilibrium potentials for Cl (0 mV) and K (−84 mV). This permits the Cl and K currents to be monitored during agonist responses (Cliff and Frizzell, 1990). Pulsing was interrupted to determine I–V relations by stepping the clamp voltage between ±100 mV at 20 mV increments (Figure 4C).

Acknowledgments

We thank Allen Korman for providing the PLJ vector, Richard Mulligan for providing ψCrip cells, Karen Murphy for assistance in sequencing, and Penny Stober for assistance in the preparation of this manuscript. This work was supported by the Cystic Fibrosis Foundation and NIH grants DK38518 (R. A. F.), DK41330 (R. A. F.), DK08260 (W. H. C.), DK34944 (L.-C. T.), DK42718 (F. S. C.), and DK39690 (F. S. C.). J. M. R. is a postdoctoral fellow for the Medical Research Council (MRC) of Canada, and L.-C. T. is a recipient of an MRC Scientist Award. J. M. W. and F. S. C. are investigators of the Howard Hughes Medical Institute.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 USC Section 1734 solely to indicate this fact.

Received September 7, 1990; revised September 13, 1990.

References

Boat, T. F., Cheny, P. W., Iyer, R., Carlson, D. M., and Polony, I. (1976). Human respiratory tract secretions: mucous glycoproteins of nonpurulent tracheobronchial secretions and sputum of patient with bronchitis and cystic fibrosis. Arch. Biochem. Biophys. *177*, 95–104.

Boucher, R. C., Stutts, M. J., Knowles, M. R., Cantley, L., and Gatzy, J. T. (1986). Na^+ transport in cystic fibrosis respiratory epithelia. Abnormal basal rate and response to adenylate cyclase activation. J. Clin. Invest. *78*, 1245–1252.

Cheng, P.-W., Boat, T. F., Cranfill, K., Yankaskas, J. R., and Boucher, R. C. (1989). Increased sulfation of glycoconjugates by cultured nasal

epithelial cells from patients with cystic fibrosis. J. Clin. Invest. *84*, 68–72.

Cliff, W. H., and Frizzell, R. A. (1990). Separate Cl^- conductances activated by cAMP and Ca^{2+} in Cl^--secreting epithelial cells. Proc. Natl. Acad. Sci. USA *87*, 4956–4960.

Cutting, G. R., Kasch, L. M., Rosenstein, B. J., Zielenski, J., Tsui, L.-C., Antonarakis, S. E., and Kazazian, H. H. (1990). A cluster of cystic fibrosis mutations in the first nucleotide-binding fold of the cystic fibrosis transmembrane conductance regulator protein. Nature *346*, 366–369.

Cystic Fibrosis Genetic Analysis Consortium (1990). Worldwide survey of the ΔF_{508} mutation—report from the Cystic Fibrosis Genetic Analysis Consortium. Am. J. Hum. Genet. *47*, 359–361.

Danos, O., and Mulligan, R. C. (1988). Safe and efficient generation of recombinant retroviruses with amphotropic and ecotropic host ranges. Proc. Natl. Acad. Sci. USA *85*, 6460–6464.

Dean, M., White, M. B., Amos, J., Gerrard, B., Stewart, C., Khaw, K.-T., and Leppert, M. (1990). Multiple mutations in highly conserved residues are found in mildly affected cystic fibrosis patients. Cell *61*, 863–870.

Feinberg, A. P., and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. *132*, 6–13.

Frizzell, R. A. (1987). Cystic fibrosis: a disease of ion channels? Trends Neurosci. *10*, 190–193.

Frizzell, R. A., Rechkemmer, G., and Shoemaker, R. L. (1986). Altered regulation of airway epithelial cell chloride channels in cystic fibrosis. Science *233*, 558–560.

Hyde, S. C., Emsley, P., Hartshorn, M. J., Mimmack, M. M., Gileadi, U., Pearce, S. R., Gallagher, M. P., Gill, D. R., Hubbard, R. E., and Higgins, C. F. (1990). Structural model of ATP-binding proteins associated with cystic fibrosis, multidrug resistance and bacterial transport. Nature *346*, 362–365.

Kerem, B.-S., Rommens, J. M., Buchanan, J. A., Markiewicz, D., Cox, T. K., Chakravarti, A., Buchwald, M., and Tsui, L.-C. (1989). Identification of the cystic fibrosis gene: genetic analysis. Science *245*, 1073–1080.

Kerem, B.-S., Zielenski, J., Markiewicz, D., Bozon, D., Gazit, E., Yahaf, J., Kennedy, D., Riordan, J. R., Collins, F. S., Rommens, J. R., and Tsui, L.-C. (1990). Identification of mutations in regions corresponding to the two putative nucleotide (ATP) binding folds of the cystic fibrosis gene. Proc. Natl. Acad. Sci. USA, in press.

Korman, A. J., Frantz, J. D., Strominger, J. L., and Mulligan, R. C. (1987). Expression of human class II major histocompatibility complex antigens using retrovirus vectors. Proc. Natl. Acad. Sci. USA *84*, 2150–2154.

Li, M., McCann, J. D., Liedtke, C. M., Nairn, A. C., Greengard, P., and Welsh, M. J. (1988). cAMP-dependent protein kinase opens chloride channels in normal but not cystic fibrosis airway epithelium. Nature *331*, 358–360.

Quinton, P. M. (1990). Cystic fibrosis: a disease in electrolyte transport. FASEB J. *4*, 2709–2717.

Riordan, J. R., Rommens, J. M., Kerem, B.-S., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J.-L., Drumm, M. L., Iannuzzi, M. C., Collins, F. S., and Tsui, L.-C. (1989). Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science *245*, 1066–1073.

Rommens, J. M., Iannuzzi, M. C., Kerem, B.-S., Drumm, M. L., Melmer, G., Dean, M., Rozmahel, R., Cole, J. L., Kennedy, D., Hidaka, M., Zsiga, M., Buchwald, M., Riordan, J. R., Tsui, L.-C., and Collins, F. S. (1989). Identification of the cystic fibrosis gene: chromosome walking and jumping. Science *245*, 1059–1065.

Sato, K., and Sato, F. (1984). Defective beta-adrenergic response of cystic fibrosis sweat glands in vivo and in vitro. J. Clin. Invest. *73*, 1763–1771.

Schoumacher, R. A., Shoemaker, R. L., Halm, D. R., Tallant, E. A., Wallace, R. W., and Frizzell, R. A. (1987). Phosphorylation fails to activate chloride channels from cystic fibrosis airway cells. Nature *330*, 752–754.

Schoumacher, R. A., Ram, J., Iannuzzi, M. C., Bradbury, N. A., Wallace, R. W., Tom Hon, C., Kelly, D. R., Schmid, S. M., Gelder, F. B., Rado, R. A., and Frizzell, R. A. (1990). A cystic fibrosis pancreatic adenocarcinoma cell line. Proc. Natl. Acad. Sci. USA *87*, 4012–4016.

Taussig, L. M. (1984). An overview. In Cystic Fibrosis, L. M. Taussig, ed. (New York: Thieme-Stratton), pp. 1–9.

Venglarik, C. J., Bridges, R. J., and Frizzell, R. A. (1990). A simple assay for agonist-regulated Cl and K conductances in salt-secreting epithelial cells. Am. J. Physiol. *259*, C358–C364.

Welsh, M. J. (1990). Abnormal regulation of ion channels in cystic fibrosis epithelia. FASEB J. *4*, 2718–2725.

White, M. B., Amos, J., Hsu, J. M. C., Gerrard, B., Finn, P., and Dean, M. (1990). A frameshift mutation in the cystic fibrosis gene. Nature *344*, 665–667.

Willumsen, N. J., and Boucher, R. C. (1989). Activation of an apical Cl-conductance by Ca^{2+} ionophores in cystic fibrosis airway epithelia. Am. J. Physiol. *256*, C226–C233.

Wilson, J. M., Johnston, D. E., Jefferson, D. M., and Mulligan, R. C. (1988). Correction of the genetic defect in hepatocytes from the Watanabe heritable hyperlipidemic rabbit. Proc. Natl. Acad. Sci. USA *85*, 4421–4425.

Wilson, J. M., Danos, O., Grossman, M., Raulet, D. H., and Mulligan, R. C. (1990a). Expression of human adenosine deaminase in mice reconstituted with retrovirus-transduced hematopoietic stem cells. Proc. Natl. Acad. Sci. USA *87*, 439–443.

Wilson, J. M., Ping, A. J., Krauss, J. C., Mayo-Bond, L., Rogers, C. E., Anderson, D. C., and Todd, R. F., III (1990b). Correction of CD18-deficient lymphocytes by retrovirus-mediated gene transfer. Science *248*, 1413–1416.

Identification of the Cystic Fibrosis Gene: Cloning and Characterization of Complementary DNA

JOHN R. RIORDAN, JOHANNA M. ROMMENS, BAT-SHEVA KEREM, NOA ALON, RICHARD ROZMAHEL, ZBYSZKO GRZELCZAK, JULIAN ZIELENSKI, SI LOK, NATASA PLAVSIC, JIA-LING CHOU, MITCHELL L. DRUMM, MICHAEL C. IANNUZZI, FRANCIS S. COLLINS, LAP-CHEE TSUI

Overlapping complementary DNA clones were isolated from epithelial cell libraries with a genomic DNA segment containing a portion of the putative cystic fibrosis (CF) locus, which is on chromosome 7. Transcripts, approximately 6500 nucleotides in size, were detectable in the tissues affected in patients with CF. The predicted protein consists of two similar motifs, each with (i) a domain having properties consistent with membrane association and (ii) a domain believed to be involved in ATP (adenosine triphosphate) binding. A deletion of three base pairs that results in the omission of a phenylalanine residue at the center of the first predicted nucleotide-binding domain was detected in CF patients.

CYSTIC FIBROSIS (CF) IS AN AUTOSOMAL RECESSIVE GENETIC disorder affecting a number of organs, including the lung airways, pancreas, and sweat glands (*1*). Abnormally high electrical potential differences have been detected across the epithelial surfaces of the CF respiratory tract, including the trachea and nasal polyps, as well as across the walls of CF sweat gland secretory coils and reabsorptive ducts (*2*). The basic defect has been associated with decreased chloride ion conductance across the apical membrane of the epithelial cells (*3*). That the defect also appeared to persist in cultured cells derived from several epithelial tissues suggested that the CF gene is expressed in these cells (*4*). More recently, patch clamp studies showed that this defect is probably due to a failure of an outwardly rectifying anion channel to respond to phosphorylation by cyclic AMP–dependent protein kinase (protein kinase A) or protein kinase C (*5*). Although progress has been made in the isolation of polypeptide components of an epithelial chloride channel that mediates conductance (*6*), their relation to the kinase-activated pathway and CF has yet to be established, and the basic biochemical defect in CF remains unknown.

Molecular cloning experiments have permitted the isolation of a large, contiguous segment of DNA spanning at least four transcribed sequences from a region thought to contain the CF locus (*7*). These sequences were initially identified on the basis of their ability to detect conserved sequences in other animal species by DNA hybridization and were subsequently characterized by RNA hybridization experiments, cDNA isolation, and direct DNA sequence analysis (*7*). Three of the transcribed regions were excluded from being the CF locus by earlier genetic or DNA sequence analyses (*7, 8*). The fourth one, as shown by genetic analysis (*9*) and DNA sequencing analysis presented below, corresponds to a portion of the CF gene locus.

Isolation of cDNA clones. Two DNA segments (E4.3 and H1.6) that detected cross-species hybridization signals (*7*) were used as probes to screen cDNA libraries made from several tissues and cell types (*10*). After screening seven different libraries, one single clone (10-1) was isolated with H1.6 from a cDNA library made from the cultured epithelial cells of the sweat glands of an unaffected (non-CF) individual (*10*).

DNA sequencing showed that 10-1 contained an insert of 920 base pairs (bp) in size and one potential, long open reading frame (ORF). Since one end of the sequence shared perfect sequence identity with H1.6, it was concluded that the cDNA clone was probably derived from this region. The DNA sequence in common was, however, only 113 bp long (Figs. 1 and 2). This sequence in fact corresponded to the first axon of the putative CF gene. The short sequence overlap thus explained the weak hybridization signals in library screening and our inability to detect transcripts in RNA gel-blot analysis. In addition, the orientation of the transcription unit was tentatively established on the basis of alignment of the genomic DNA sequence with the presumptive ORF of 10-1.

Since the corresponding transcript was estimated to be about 6500 nucleotides in length by RNA gel-blot hybridization experiments, further cDNA library screening was required in order to clone the remainder of the coding region. As a result of several successive screenings with cDNA libraries generated from the colon carcinoma cell line T84, normal and CF sweat gland cells, pancreas,

J. R. Riordan, N. Alon, and Z. Grzelczak are in the Department of Biochemistry, Research Institute, The Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. J. M. Rommens, B. Kerem, R. Rozmahel, J. Zielenski, S. Lok, N. Plavsic, J.-L. Chou, and L.-C. Tsui are at the Department of Genetics, Research Institute, The Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. M. L. Drumm, M. C. Iannuzzi, and F. S. Collins are at the Howard Hughes Medical Institute and Departments of Internal Medicine and Human Genetics, University of Michigan, Ann Arbor, MI 48109. J. R. Riordan is also a member of the Departments of Biochemistry and Clinical Biochemistry, and L.-C. Tsui is also a member of the Departments of Medical Genetics and Medical Biophysics, University of Toronto, Toronto, Ontario M5S 1A8, Canada.

and adult lungs, 18 additional clones were isolated (Fig. 1). DNA sequence analysis revealed that none of these cDNA clones corresponded to the length of the observed transcript, but it was possible to derive a consensus sequence based on overlapping regions. Further cDNA clones corresponding to the 5′ and 3′ ends of the transcript were derived from 5′ and 3′ primer-extension experiments (Fig. 1). Together, these clones span about 6.1 kb and contain an ORF capable of encoding a protein of 1480 amino acids (Fig. 2).

It was unusual that most of the cDNA clones isolated here contained sequence insertions at various locations (Fig. 1). While many of these extra sequences corresponded to intron regions reverse-transcribed during the construction of the cDNA, as revealed on alignment with genomic DNA sequences, the identities of several others were uncertain because they did not align with sequences at the corresponding exon-intron junctions, namely, the sequences at the 5′ ends of clones 13a and T16-1 and at the 5′ and 3′ ends of T11, and the insertions between exons 3 and 4 in 13a and between exons 10 and 11 in T16-4.5 (legend to Fig. 1). More puzzling were the sequences corresponding to the reverse complement of exon 6 at the 5′ end of 11a and the insertion of a segment of a bacterial transposon in clone C16-1; none of these could be explained by mRNA processing errors.

In that the number of recombinant cDNA clones for the putative CF gene detected in the library screening was much less than would have been expected from the abundance of transcripts estimated from RNA hybridization experiments, it seemed probable that the clones that contained aberrant structures were preferentially retained while the proper clones were lost during propagation. Consistent with this interpretation, poor growth was observed for most of our recombinant clones isolated, regardless of the vector used.

RNA analysis. To visualize the transcript of the putative CF gene, we used RNA gel-blot hybridization with the 10-1 cDNA as the probe (Fig. 3). The analysis revealed a prominent band, about 6.5 kb in size, in T84 cells. Identical results were obtained with other cDNA clones as probes. Similar, strong hybridization signals were also detected in pancreas and primary cultures of cells from nasal polyps, suggesting that the mature mRNA of the putative CF gene is about 6.5 kb. Minor hybridization signals, probably representing degradation products, were detected at the lower size ranges, but they varied between different experiments. On the basis of the hybridization band intensity and comparison with those detected for other transcripts under identical experimental conditions, it was estimated that the putative CF gene transcripts constituted about 0.01 percent of total mRNA in T84 cells.

Additional tissues were analyzed by RNA gel-blot hybridization in an attempt to correlate the expression pattern of the putative CF gene and the pathology of CF. Transcripts, all of identical size, were found in lung, colon, sweat glands (cultured epithelial cells), placenta, liver, and parotid gland, but the signal in these tissues was generally weaker than that detected in the pancreas and nasal polyps (Fig. 3). Intensity varied among different preparations; for example, hybridization in kidney was not detectable in the preparation shown in Fig. 3 but was clearly discernible subsequently. Transcripts were not detected in the brain or adrenal gland, nor in skin fibroblast and lymphoblast cell lines.

Thus, expression of the putative CF gene appeared to occur in many of the tissues examined, with higher levels in those tissues severely affected in CF. While this epithelial tissue–specific expression pattern is in good agreement with the disease pathology, no significant difference was detected in the amount or size of transcripts from CF and control tissues (Fig. 3), consistent with the assumption that CF mutations are subtle changes at the nucleotide level.

Characterization of cDNA clones. As indicated above, a contig-

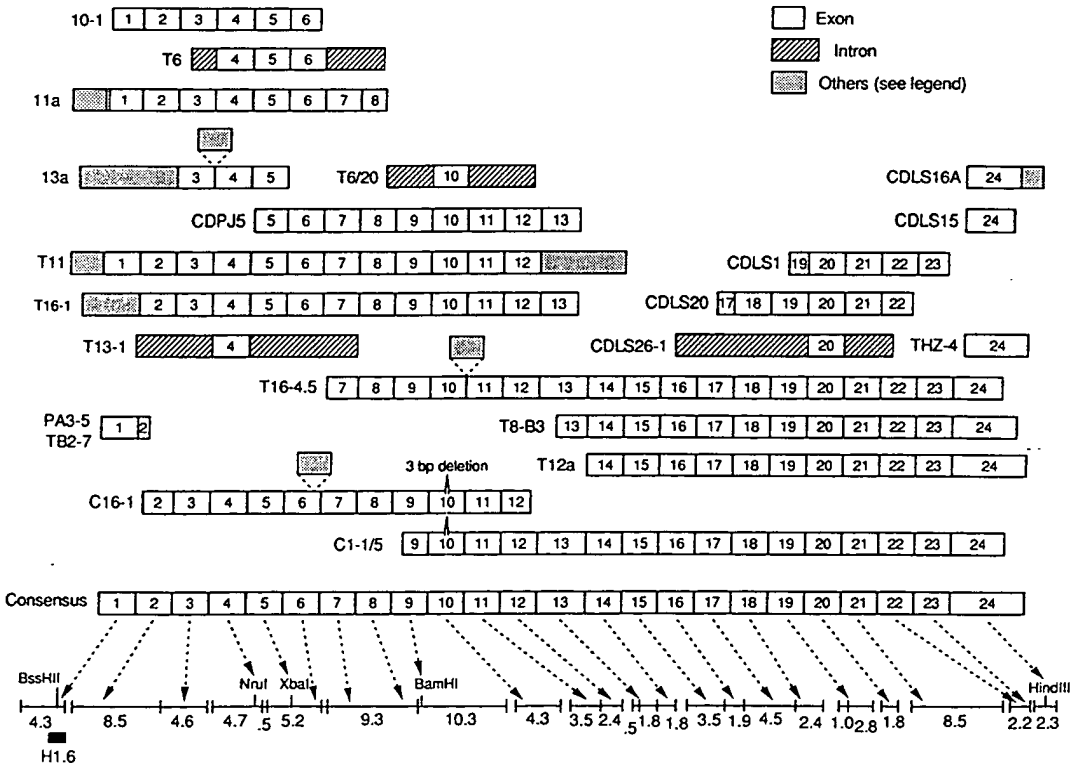
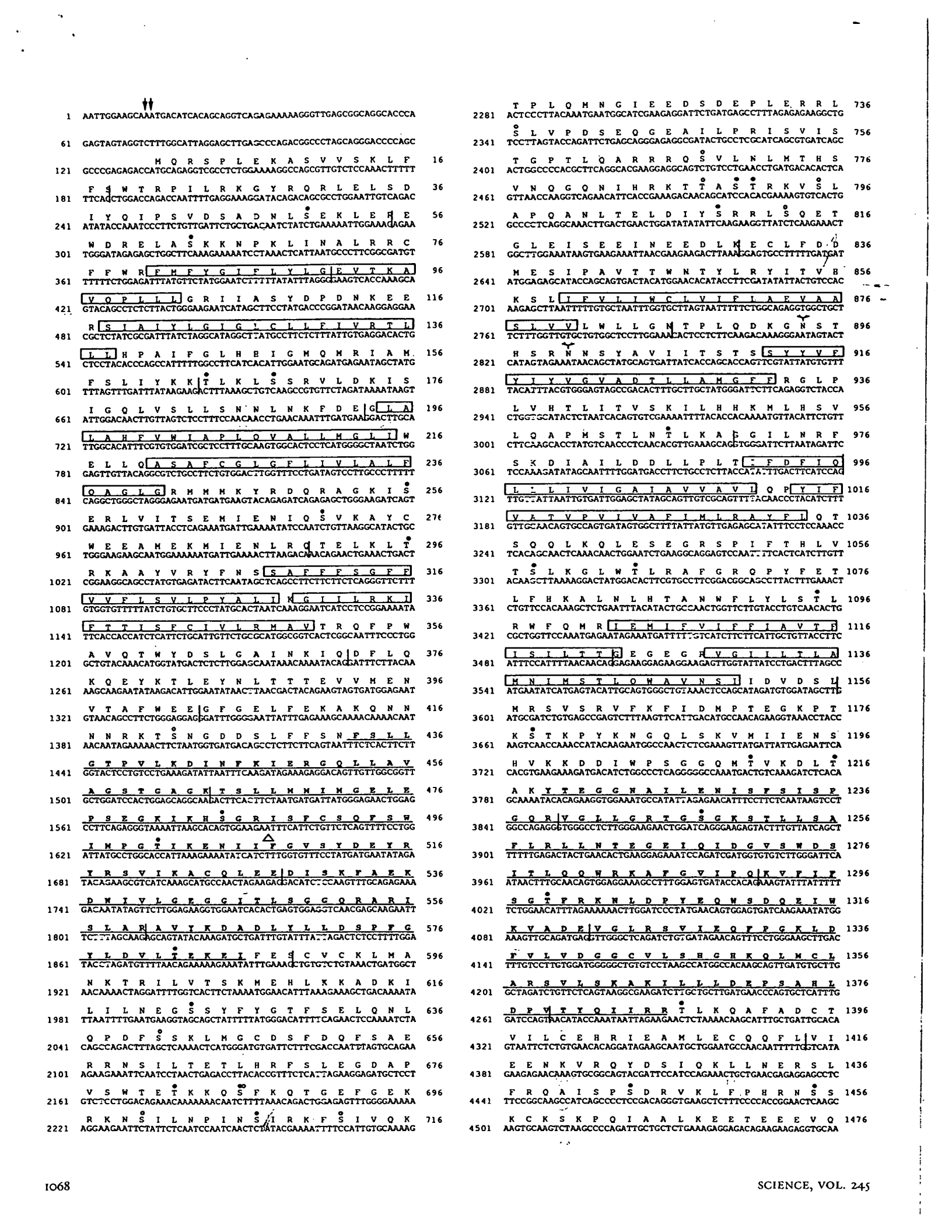


Fig. 1. Overlapping cDNA clones aligned with genomic DNA fragments. The cDNA clones are represented by open boxes with exons indicated. The corresponding genomic Eco RI fragments are schematically presented on the bottom, with lengths in kilobases. The hatched boxes denote intron sequences, and stippled boxes represent other sequences as outlined below. The filled box in the lower left is the position of the clone H1.6, which was used to isolate the first cDNA clone 10-1 from a normal (N) sweat gland library (*10*). The definitive restriction sites used for the alignment of cDNA and genomic fragments are indicated. Clones T6, T6/20, T11, T16-1, T13-1, T16-4.5, T8-B3, and T12a were isolated sequentially from the T84 cell library (*10*). Clones isolated from the human lung cDNA library (*10*) are designated with the prefix CDL. CDPJ5 is derived from a pancreas library (*10*). The CF sweat gland cDNA clones, C16-1 and C1-1/5, together cover all but exon 1 and a portion of the 3′ untranslated region. Both clones revealed a 3-bp deletion in exon 10. Clones that contain intron sequences are CDLS26-1, T6/20, and T13-1. Clones T11, T16-4.5, CDLS16A, 11a, and 13a contain extraneous sequences of unknown origin at positions indicated. Clone C16-1 contains a short insertion corresponding to a portion of the γ transposon of *E. coli*. Both PA3-5 and TB2-7 are 5′ extension clones generated from pancreas and T84 RNA by the anchored PCR technique (*12*), respectively. THZ-4 is a 3′ extension clone obtained from T84 RNA. Both T12a and THZ-4 contain a polyadenylation signal and a poly(A)$^+$ tail.



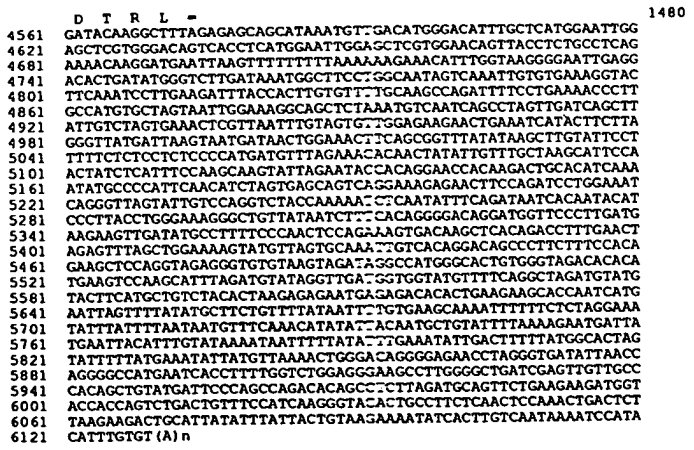


Fig. 2. Nucleotide sequence of cDNA encoding the CF transmembrane conductance regulator together with the deduced amino acid sequence. DNA sequencing was performed by the dideoxy chain termination method (*34*) with ^{35}S-labeled nucleotides or by the Dupont Genesis2000 automatic DNA sequencer. Numbers on the left of columns indicate base positions and numbers on the right amino acid residue positions. The first base position corresponds to the first nucleotide in the 5′ extension clone PA3-5, which is one nucleotide longer than TB2-7 (*12*). The 3′ end and the noncoding sequence are shown above [nucleotides 4561 to 6129 plus the poly(A)$^+$ tail]. Arrows indicate position of transcription initiation site by primer extention analysis (*11*). Nucleotide 6129 is followed by a poly(A) tract. Positions of exon junctions are indicated by vertical lines. Potential membrane-spanning segments ascertained with the use of the algorithm of Eisenberg *et al.* (*35*) are enclosed in boxes. Amino acids comprising putative ATP-binding folds are underlined. Possible sites of phosphorylation (*21*) by protein kinases A or C are indicated by open and closed circles, respectively. The open triangle indicates the position at which 3 bp are deleted in CF. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

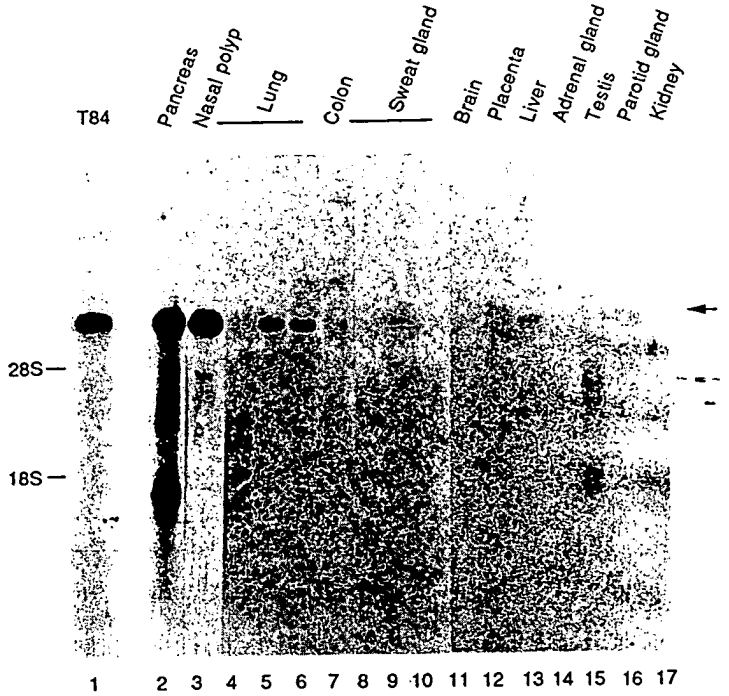


Fig. 3. RNA gel-blot analysis. Hybridization by the cDNA clone 10-1 to a 6.5-kb transcript is shown in the tissues indicated. RNA samples were prepared from cells and tissue samples obtained from surgical pathology or at autopsy according to the methods described in (*10*). Total RNA (10 μg) from each tissue and 1 μg of poly(A)$^+$ RNA from T84 cells were separated on formaldehyde gels and transferred onto nylon membranes (Zetaprobe, Bio-Rad), which were hybridized with DNA probes labeled to high specific activity by the random priming method (*36, 37*). The positions of the 28*S* and 18*S* rRNA bands are indicated.

uous coding region of the CF locus could be deduced from overlapping cDNA clones. Since most of the cDNA clones were apparently derived from unprocessed transcripts, further studies were performed to ensure the authenticity of the consensus sequence. Each cDNA clone was first tested for chromosome localization by hybridization analysis with a human-hamster somatic cell hybrid containing a single human chromosome 7 and by pulsed field gel electrophoresis (7). The ones that did not map to the correct region on chromosome 7 were not pursued. Fine restriction enzyme mapping was then performed for each clone. While overlapping regions were clearly identified for most of the clones, many contained single copy, additional regions not readily recognizable by restriction enzyme analysis.

The cDNA was further characterized in gel hybridization experiments with genomic DNA. Five to six different restriction fragments could be detected with the 10-1 cDNA in Eco RI– or Hind III–digested total human DNA and a similar number of fragments with several other cDNA clones, suggesting the presence of multiple exons for the putative CF gene. The hybridization studies also identified the cDNA clones with unprocessed intron sequences when they showed preferential hybridization to a smaller subset of genomic DNA fragments with relatively greater intensities. For the confirmed cDNA clones, their corresponding genomic DNA segments were isolated (7) and the exons and exon-intron boundaries were sequenced. In all, 24 exons were identified (Fig. 2). Physical mapping experiments (7) showed that the gene locus spanned about 250 kb.

The 5′ terminus of the transcript was determined by primer extension (*11*). A modified polymerase chain reaction, anchored PCR (*12*), was also used to facilitate cloning of the 5′ end sequences.

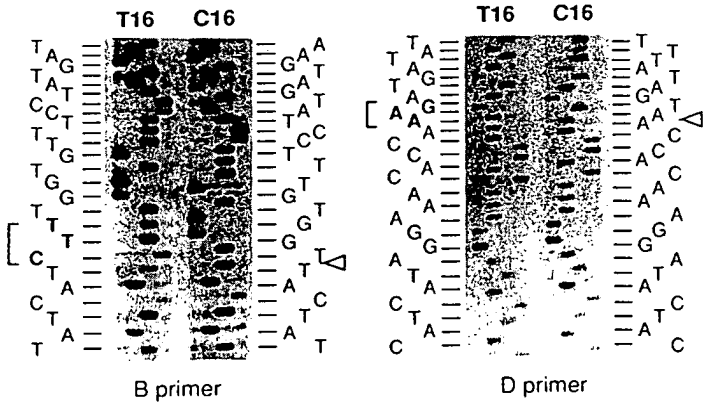


Fig. 4. DNA sequence around the ΔF_{508} deletion. The normal sequence from base position 1627 to 1651 (from cDNA T16-1) is shown beside the CF sequence (from cDNA C16-1). The left panel shows the sequences from the coding strands obtained with the B primer (5′-GTTTTCCTGGATTATGCCTGGGCAC-3′) and the right panel those from the opposite strand with the D primer (5′-GTTGGCATGCTTTGATGACGCTTC-3′). The brackets indicate the three nucleotides in the normal that are absent in CF (arrowheads). Sequencing was performed as described in (*34*).

Two independent 5′ extension clones, one from pancreas and the other from T84 RNA, were characterized by DNA sequencing and differed by only 1 base in length, thus establishing the most probable initiation site for the transcript (Fig. 2). Since the initial cDNA clones did not contain a poly(A)$^+$ tail indicative of the end of a mRNA, anchored PCR was also applied to the 3′ end of the transcript (*12*). The results derived from the use of several different 3′-extending oligonucleotides were consistent with the interpreta-

tion that the end of the transcript was about 1.2 kb downstream of the Hind III site at nucleotide position 5027 (Fig. 2).

The complete cDNA sequence spans 6129 base pairs excluding the poly(A)$^+$ tail at the end of the 3′ untranslated region and it contains an ORF capable of encoding a polypeptide of 1480 amino acids (Fig. 2). An ATG (AUG) triplet is present at the beginning of this ORF (base position 133–135). Since the nucleotide sequence surrounding this codon (5′-AGACCAUGCA-3′) has the proposed features of the consensus sequence (CC)$^{A}_{G}$CCAUGG(G) of a eukaryotic translation initiation site (*13*), with a highly conserved A at the −3 position, it is highly probable that this AUG corresponds to the first methionine codon for the putative polypeptide.

Detection of mutation. A comparison between the cDNA sequences derived from CF and unaffected (N) individuals was next conducted. Two clones, C16-1 and C1-1/5, were derived from a CF sweat gland cDNA library and together they spanned almost the entire coding region. The most striking difference between CF and N sequences was a 3-bp deletion (Fig. 4), which would result in a loss of a phenylalanine residue (position 508) in the predicted CF polypeptide. This deletion (ΔF_{508}) was detected in both CF clones. To exclude the possibility that this difference was due to a cloning artifact, sequence-specific oligonucleotides were used to screen DNA samples from CF families. Specific hybridization could be observed for each oligonucleotide probe with genomic DNA amplified by PCR, confirming the presence of corresponding genomic DNA sequences (*9*). Furthermore, the oligonucleotide specific for the 3-bp deletion hybridized to 68 percent of chromosomes carrying a CF mutation but not to any of the normal chromosomes (0/198), an indication that a silent sequence polymorphism was unlikely. Sequence differences found elsewhere among the different cDNA clones probably represented sequence polymorphisms or cDNA cloning artifacts (*14*).

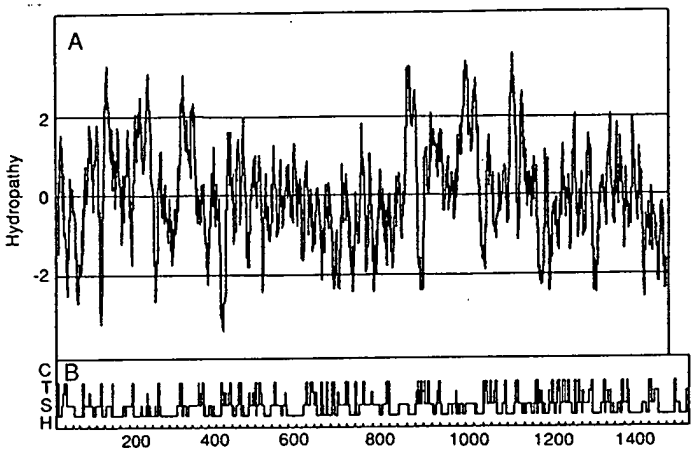


Fig. 5. Hydropathy profile and predicted secondary structures of the CFTR. (**A**) The mean hydropathy index determined according to Kyte and Doolittle (*19*) of nine-residue peptides is plotted against the amino acid number. (**B**) The corresponding positions of features of secondary structure predicted according to Garnier *et al.* (*19*). C, coil; T, turn; S, sheet; H, helix.

Predicted protein structure. Analysis of the sequence of the overlapping cDNA clones (Fig. 2) predicted a polypeptide of 1480 amino acids with a molecular mass of 168,138 daltons. The most characteristic feature of the predicted protein is the presence of two repeated motifs, each of which consists of a domain capable of spanning the membrane several times and sequences resembling consensus nucleotide (ATP)-binding folds (NBF's) (*15*) (Figs. 5 and 6). These characteristics are remarkably similar to those of the mammalian multidrug resistance P-glycoprotein (*16*) and a number of other membrane-associated proteins (as discussed below), suggesting that the predicted CF gene product is likely to be involved in the transport of substances (ions) across the membrane and is probably a member of a membrane protein superfamily (*17*). For the convenience of future discussion and to avoid confusion with the previously named CF protein and CF factor (*18*), we will call the

```
                                                                  *
CFTR (N)   FSLLGTPVLKDINFKIERGQLLAVAGSTGAGKTSLLMMIMG  ISFCSQFSWIMPGTIK-ENIIFGVSYD  GEGGITLSGGQRARISLARAVYKDADLYLLDSPFGYLDVLTEK
CFTR (C)   YTEGGNAILENISFSISPGQRVGLLGRTGSGKSTLLSAFLR  DSITLQQWRKAFGVIPQKVFIFSGTFR  VDGGCVLSHGHKQLMCLARSVLSKAKILLLDEPSAHLDPVTYQ
hmdr1 (N)  PSRKEVKILKGLNLKVQSGQTVALVGNSGCGKSTTVQLMQR  IGVVSQEPVLFATTI-AENIRYGRENV  GERGAQLSGGQKQRIAIARALVRNPKILLLDEATSALDTESEA
hmdr1 (C)  PTRPDIPVLQGLSLEVKKGQTLALVGSSGCGKSTVVQLLER  LGIVSQEPILFDCSI-AENIAYGDNSR  GDKGTLLSGGQKQRIAIARALVRQPHILLLDEATSALDTESEK
mmdr1 (N)  PSRSEVQILKGLNLKVKSGQTVALVGNSGCGKSTTVQLMQR  IGVVSQEPVLFATTI-AENIRYGREDV  GERGAQLSGGQKQRIAIARALVRNPKILLLDEATSALDTESEA
mmdr1 (C)  PTRPNIPVLQGLSLEVKKGQTLALVGSSGCGKSTVVQLLER  LGEVSQEPILFDCSI-AENIAYGDNSR  GDKGTQLSGGQKQRIAIARALVRQPHILLLDEATSALDTESEK
mmdr2 (N)  PSRANIKILKGLNLKVKSGQTVALVGNSGCGKSTTVQLLQR  IGVVSQEPVLSFTTI-AENIRYGRGNV  GDRGAQLSGGQKQRIAIARALVRNPKILLLDEATSALDTESEA
mmdr2 (C)  PTRANVPVLQGLSLEVKKGQTLALVGSSGCGKSTVVQLLER  LGIVSQEPILFDCSI-AENIAYGDNSR  GDKGTQLSGGQKQRIAIARALIRQPRVLLLDEATSALDTESEK
pfmdr (N)  DTRKDVEIYKDLSFTLLKEGKTYAFVGESGCGKSTILKLIE  IGVVSQDPLLFSNSI-KNNIKYSLYSL  GSNASKLSGGQKQRISIARAIMRNPKILILDEATSSLDNKSEY
pfmdr (C)  ISRPNVPIYKNLSFTCDSKKTTAIVGETGSGKSTFMNLLLR  FSIVSQEPMLFNMSI-YENIKFGREDA  PYGKS-LSGGQKQRIAIARALLREPKILLLDEATSSLDSNSEK
STE6 (N)   PSRPSEAVLKNVSLNFSAGQFTFIVGKSGSGKSTLSNLLLR  ITVVEQRCTLFNDTL-RKNILLGSTDS  GTGGVTLSGGQQQRVAIARAFIRDTPILFLDEAVSALDIVHRN
STE6 (C)   PSAPTAFVYKNMNFDMFCGQTLGIIGESGTGKSTLVLLLTK  ISVVEQKPLLFNGTI-RDNLTYGLQDE  RIDTTLLSGGQAQRLCIARALLRKSKILILDECTSALDSVSSS
hlyB       YKPDSPVILDNINISIKQGEVIGIVGRSGSGKSTLIKLIQR  VGVVLQDNVLLNRSI-IDNISLAPGMS  GEQGAGLSGGQRQRIAIARALVNNPKILIFDEATSALDYASEH
White      IPAPRKHLLKNVCGVAYPGELLAVMGSSGAGKTTLLNALAF  RCAYVQQDDLFIGLIAREHLIFQAMVR  PGRVKGLSGGERKRLAFASEALTDPPLLICDEPTSGLDSFTAH
MbpX       KSLGNLKILDRVSLYVPKFSLIALLGPSGSGKSSLLRILAG  MSFVFQHYALFKHMTVYENISFGLRLR  FEYPAQLSGGQKQRVALARSLAIQPDLLL-DEPFGALDGELRR
BtuD       QDVAESTRLGPLSGEVRAGRILHLVGPNGAGKSTLLARIAG  YLSQQQTPPFATPVWHYLTLHQHDKTR  GRSTNQLSGGEWQRVRLAAVVLQITLLLLDEPMNSLDVAQQSA
PstB       FYYGKFHALKNINLDTAKNQVTAFIGPSGCGKSTLLRTFNK  VGMVFQKPTPFPMSI-YDNIAFGVRLF  HQSGYSLSGGQQQRLCIARGIAIRPEVLLLDEPCSALDPISTG
hisP       RRYGGHEVLKGVSLQARAGDVISIIGSSGSGKSTFLRCINF  GIMVFQHFNLWSHMTVLENVMEAPIQV  GKYPVHLSGGQQQRVSIARALAMEPDVLLFDEPTSALDPELVG
malK       KAWGEVVVSKDINIDIHEGEFVVFVGPSGCGKSTLLRMIAG  VGMVFQSYALYPHLSVAENMSFGLKPA  DRKPKALSGGQRQRVAIGRTLVAEPSVFLLDEPLSNLDAALRV
oppD       TPDGDVTAVNDLNFTLRAGETLGIVGESGSGKSQTAFALMG  ISMIFQDPMTSLNPYMRVGEQLMEVLM  KMYPHEFSGGMRQRVMIAMALLCRPKLLIADEPTTALDVTVQA
oppF       QPPKTLKAVDGVTLRLYEGETLGVVGESGCGKSTFARAIIG  IQMIFQDPLASLNPRMTIGEIIAEPLR  NRYPHEFSGGQCQRIGIARALILEPKLIICDDAVSALDVSIQA
RbsA (N)   KAVPGVKALSGAALNVYPGRVMALVGENGAGKSTMMKVLTG  AGIIHQELNLIPQLTIAENIFLGREFV  DKLVGDLSIGDQQMVEIAKVLSFESKVIIMDEPTCALIDTETE
RbsA (C)   VDNLCGPGVNDVSFTLRKGEILGVSGLMGAGRTELMKVLYG  ISEDRKRDGLVLGMSVKENMSLTALRY  EQAIGLLSGGNQQKVAIARGLMTRPKVLILDEPTPGVDVGAKK
UvrA       LTGARGNNLKDVTLTLPVGLFTCITGVSGSGKSTLINDTLF  TYTGVFTPVRELFAGVPESRARGYTPG  GQSATTLSGGEAQRVKLARELSKRGLYILDEPTTGLHFADIQQ
NodI       KSYGGKIVVNDLSFTIAAGECFGLLGPNGAGKSTIIRMILG  IGIVSQEDNLDLEFTVRENLLVYGRYF  NTRVADLSGGMKRRLTLAGALINDPQLLILDEPTTGLDPHARH
FtsE       AYLGGRQALQGVTFHMQPGEMAFLTGHSGAGKSTLLKLICG  IGMIFQDHHLLMDRTVYDNVAIPLIA  KNFPIQLSGGEQQRVGIARAVVNKPAVLLADEPTGNLDDALSE
```

Fig. 6. Alignment of the three most conserved segments of the amino acid sequences (single letter code) of the extended NBF's of CFTR with comparable regions of other proteins. These three segments consist of residues 433 to 473, 488 to 513, and 542 to 584 of the amino-terminal (N) half and 1219 to 1259, 1277 to 1302, and 1340 to 1382 of the carboxyl-terminal (C) half of CFTR. The heavy overlining points out the regions of greatest similarity. The star indicates the position corresponding to the phenylalanine that is deleted in CF. Additional general homology can be seen even with the introduction of very few gaps. The other sequences are of proteins involved in multidrug resistance in human (hmdr1), mouse (mmdr 1 and 2) (*16*), and *Plasmodium falciparum* (pfmdr) (*38*); the α-factor pheromone export system of yeast (STE6) (*39*); the hemolysin (hlyB) system of *E. coli* (*22*); screening of eye pigments in *Drosophila* (White) (*23*); an unknown liverwort chloroplast function (Mbpx) (*25*); vitamin B12 transport in *E. coli* (BtuD) (*24*); phosphate transport in *E. coli* (PstB) (*40*); histidine transport in *Salmonella typhimurium* (hisP) (*41*); maltose transport in *E. coli* (malK) (*42*); oligopeptide transport in *S. typhimurium* (oppD and oppF) (*43*); ribose transport in *E. coli* (RbsA) (*44*). UvrA is one component of an *E. coli* DNA repair system (*45*); NodI is a gene product involved in nodulation in *Rhizobium* (*46*); FtsE is a protein that contributes to the regulation of cell division (*47*). In addition to these proteins that contain this long NBF, there is a large number of others that contain the two short nucleotide binding motifs A and B initially pointed out by Walker *et al.* (*48*). Further, there are other proteins containing only motif A or B (*49*).

putative CF gene product the cystic fibrosis transmembrane conductance regulator (CFTR).

Each of the predicted membrane-associated regions of CFTR consists of six hydrophobic segments capable of spanning a lipid bilayer (*19*), which are followed by a large hydrophilic region containing the NBF's (Fig. 5). On the basis of sequence alignment with other nucleotide-binding proteins, each of the putative NBF's in CFTR comprises at least 150 residues (Fig. 6). The single residue deletion (ΔF_{508}) detected in most of the CF patients is in the first NBF, between the two most highly conserved segments within this sequence. The amino acid sequence identity between the region surrounding the ΔF_{508} mutation and the corresponding regions of several other proteins suggests that this region is of functional importance (Fig. 6). A hydrophobic amino acid, usually one with an aromatic side chain, is present in most of these proteins at the position corresponding to Phe^{508} of CFTR.

Despite the overall symmetry in the two-motif structure of the protein and the sequence conservation of the NBF's, sequence identity between the two motifs of the predicted CFTR protein is modest. The strongest identity is between sequences at the carboxyl ends of the NBF's. Of the 66 residues aligned within these regions, 27 percent are identical and 11 percent are functionally similar. The overall, weak internal sequence identity is in contrast to the much higher degree (>70 percent) in P-glycoprotein for which a sequence duplication hypothesis has been proposed (*16*). The lack of conservation in the relative positions of the exon-intron boundaries in the CF gene also argues against recent exon duplication as a mechanism in the evolution of this gene (Fig. 2).

Since there is apparently no signal-peptide sequence at the amino terminus of CFTR (Fig. 7), the highly charged hydrophilic segment preceding the first transmembrane sequence is probably oriented in the cytoplasm. Each of the two sets of hydrophobic helices are expected to form three traversing loops across the membrane and little of the sequence of the entire protein is expected to be exposed to the exterior surface, except the region between transmembrane segments 7 and 8. It is of interest that the latter region contains two potential sites for N-linked glycosylation (*20*).

A highly charged cytoplasmic domain can be identified in the middle of the predicted CFTR polypeptide, linking the two halves of the protein. This domain, named the R domain, is operationally defined by a single large exon in which 69 of the 241 amino acids are polar residues arranged in alternating clusters of positive and negative charges. Moreover, nine of the ten sites at which there are consensus sequences for phosphorylation by protein kinase A and seven of the potential substrate sites for protein kinase C found in CFTR are located in this exon (*21*).

Properties of CFTR could be further derived from comparison to other membrane-associated proteins (Fig. 6). In addition to the overall structural similarity with P-glycoproteins, each of the two predicted motifs in CFTR shows resemblance to the single motif structure of hemolysin B of *Escherichia coli* (*22*) and the product of the White gene of *Drosophila* (*23*). These proteins are involved in the transport of the lytic peptide of the hemolysin system and of eye pigment molecules, respectively. The vitamin B12 transport system of *E. coli*, BtuD (*24*), and MbpX (*25*), which is a liverwort chloroplast gene product whose function is unknown, also have a similar structural motif. Further, CFTR shares structural similarity with several of the periplasmic solute transport systems of Gram-negative bacteria, where the transmembrane region and the ATP-binding folds are contained in separate proteins that function in concert with a third substrate-binding polypeptide (*26*).

The overall structural arrangement of the transmembrane domains in CFTR is similar to several cation channel proteins (*27*) and some cation-translocating adenosine triphosphatases (ATPases) (*28*)

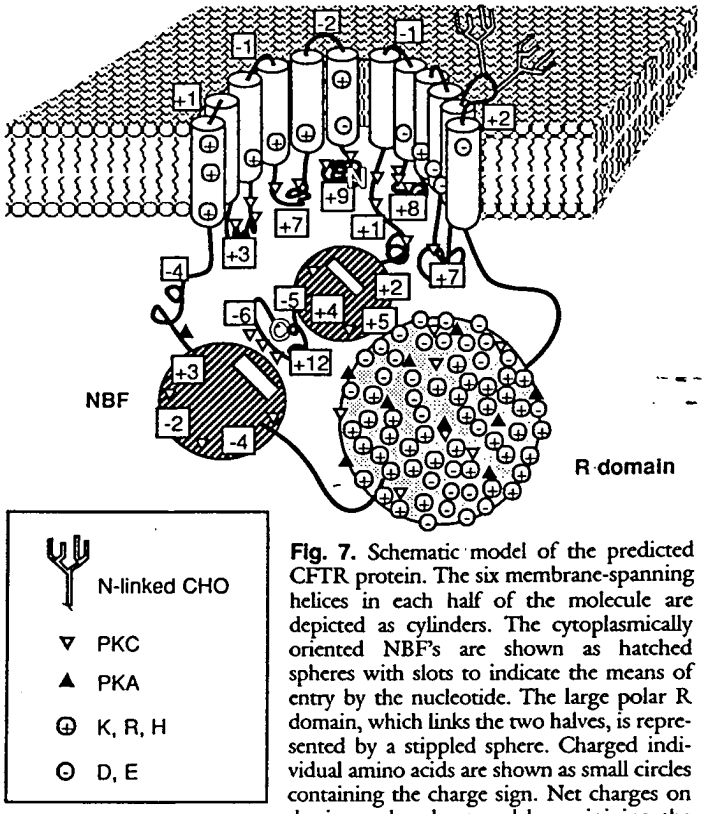


Fig. 7. Schematic model of the predicted CFTR protein. The six membrane-spanning helices in each half of the molecule are depicted as cylinders. The cytoplasmically oriented NBF's are shown as hatched spheres with slots to indicate the means of entry by the nucleotide. The large polar R domain, which links the two halves, is represented by a stippled sphere. Charged individual amino acids are shown as small circles containing the charge sign. Net charges on the internal and external loops joining the membrane cylinders and on regions of the NBF's are contained in open squares. Potential sites for phosphorylation by protein kinases A or C (PKA or PKC) and N-glycosylation (N-linked CHO) are as indicated. K, Lys; R, Arg; H, His; D, Asp; and E, Glu.

as well as the recently described adenylate cyclase of bovine brain (*29*). Short regions of sequence identity have also been detected between the putative transmembrane regions of CFTR and other membrane-spanning proteins (*30*). In addition, a sequence of 18 amino acids situated approximately 50 residues from the carboxyl terminus of CFTR shows some identity (12/18) with the raf serine-threonine kinase proto-oncogene product of *Xenopus laevis* (*31*).

Finally, a sequence identity (10 of 13 amino acid residues) has been noted between a hydrophilic segment (position 701 to 713) within the highly charged R domain of CFTR and a region immediately preceding the first transmembrane loop of the sodium channels in both rat brain and eel (*32*). This feature of CFTR is not shared with the topologically closely related P-glycoprotein; the 241–amino acid linking peptide is apparently the major difference between the two proteins.

Relevance to the CF anion transport defect. In view of the genetic data of Kerem *et al.* (*9*) and the tissue specificity and predicted properties of the CFTR protein, it is reasonable to conclude that CFTR is directly responsible for CF. It remains unclear, however, how CFTR is involved in the regulation of ion conductance across the apical membrane of epithelial cells.

It is possible that CFTR serves as an ion channel itself. For example, 10 of the 12 putative transmembrane regions contain one or more amino acids with charged side chains (Fig. 7), a property similar to that of the brain sodium channel and the γ-aminobutyric acid (GABA) receptor chloride channel subunits, where charged residues are present in four of the six, and three of the four, respective membrane-associated domains per subunit or repeat unit (*32, 33*). The amphipathic nature of these transmembrane segments is believed to contribute to the channel-forming capacity of these molecules. In contrast, the closely related P-glycoprotein, which is

not believed to conduct ions, has only two charged residues in all 12 transmembrane domains. Alternatively, CFTR may not be an ion channel but instead it may serve to regulate ion channel activities. In support of the latter possibility, none of the recently purified polypeptides (from trachea and kidney) that are capable of reconstituting chloride channels in lipid membranes (*6*) appear to be CFTR, judged on the basis of molecular mass.

In any case, the presence of ATP-binding domains in CFTR suggests that ATP hydrolysis is directly involved and required for the transport function. The high density of phosphorylation sites for protein kinases A and C and the clusters of charged residues in the R domain may both serve to regulate this activity. The deletion of Phe^{508} in the NBF may prevent proper binding of ATP or the conformational change required for normal CFTR activity, consequently resulting in the observed insensitivity to activation by protein kinase A– or protein kinase C–mediated phosphorylation of the CF apical chloride conductance pathway (*5*). Since the predicted structure of CFTR contains several conserved domains and belongs to a family of proteins, most of which function as parts of multicomponent molecular systems (*15*), the CFTR protein may also participate in epithelial cell functions not related to ion transport.

To understand the basic defect in CF, it is necessary to determine the precise role of Phe^{508} in the regulation of ion transport and to understand the mechanism that leads to the pathophysiology of the disease. With the CF gene (that is, the cDNA) now isolated, it should be possible to elucidate the control of ion transport pathways in epithelial cells in general. Knowledge gained from study of the CF gene product (CFTR), both the normal and mutant forms, will provide a molecular basis for the development of improved means of treatment of the disease.

REFERENCES AND NOTES

1. T. F. Boat, M. J. Welsh, A. L. Beaudet, in *The Metabolic Basis of Inherited Disease*, C. L. Scriver, A. L. Beaudet, W. S. Sly, D. Valle, Eds. (McGraw-Hill, New York, ed. 6, 1989), pp. 2649–2680.
2. M. R. Knowles *et al.*, *Science* **221**, 1067 (1983); M. R. Knowles, J. Gatzy, R. Boucher, *J. Clin. Invest.* **71**, 1410 (1983); P. M. Quinton, *Nature* **301**, 421 (1983); K. Sato, *Am. J. Physiol.* **247**, R646 (1984).
3. J. H. Widdicombe, M. J. Welsh, W. E. Finkbeiner, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 6167 (1985); P. M. Quinton and J. Bijman, *N. Engl. J. Med.* **308**, 1185 (1983).
4. J. R. Yankaskas, C. U. Cotton, M. R. Knowles, J. T. Gatzy, R. C. Boucher, *Am. Rev. Resp. Dis.* **132**, 1281 (1985); P. S. Pedersen, E. H. Larsen, B. Hainau, N. J. Brandt, *Med. Sci. Res.* **15**, 1009 (1987); T. J. Jensen and J. R. Riordan, unpublished results.
5. M. J. Welsh, *Science* **232**, 1648 (1986); R. A. Frizzell, G. Rechkemmer, R. L. Shoemaker, *ibid.* **233**, 558 (1986); M. J. Welsh and C. M. Liedtke, *Nature* **322**, 467 (1986); R. A. Schoumacher *et al.*, *ibid.* **330**, 752 (1987); M. Li *et al.*, *ibid.* **331**, 358 (1988); T.-C. Hwang *et al.*, *Science* **244**, 1351 (1989); M. Li *et al.*, *ibid.*, p. 1353.
6. D. W. Landry *et al.*, *Science* **244**, 1469 (1989).
7. J. M. Rommens *et al.*, *ibid.* **245**, 1059 (1989).
8. M. Farrall *et al.*, *Am. J. Hum. Genet.* **43**, 471 (1988).
9. B. Kerem *et al.*, *Science* **245**, 1073 (1989).
10. The cDNA libraries from cultured epithelial cells were prepared as follows: sweat gland cells derived from a non-CF individual and from a CF patient were grown to first passage as described [G. Collie, M. Buchwald, P. Harper, J. R. Riordan, *In Vitro Cell. Dev. Biol.* **21**, 592 (1985)]. The presence in these cells of an outwardly rectifying Cl^- channel was confirmed (J. A. Tabcharani, T. J. Jensen, J. R. Riordan, J. W. Hanrahan, *J. Membrane Biol.*, in press), but the CF cells were insensitive to activation by cyclic AMP [T. J. Jensen, J. W. Hanrahan, J. A. Tabcharani, M. Buchwald, J. R. Riordan, *Pediatric Pulmonol. Suppl.* **2**, 100 (1988)]. Polyadenylated RNA was isolated [J. M. Chirgwin, A. E. Przybyla, R. J. Macdonald, W. J. Rutter, *Biochemistry* **18**, 5294 (1979); H. Aviv and P. Leder, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 1408 (1972)] and used as template for the synthesis of cDNA according to U. Gubler and B. Hoffman [*Gene* **25**, 263 (1983)]. After methylation of internal Eco RI sites, ends were made flush with T4 DNA polymerase, and phosphorylated Eco RI linkers were added to the cDNA. After digestion with Eco RI and removal of excess linkers, the cDNA products were ligated into the Eco RI site of λ ZAP (Stratagene, San Diego, CA). The same procedures were used to construct a library from RNA isolated from preconfluent cultures of the T84 colon carcinoma cell line [K. Dharmsathaphorn, J. A. McRoberts, K. G. Mandel, L. D. Tisdale, H. Masui, *Am. J. Physiol.* **246**, G204 (1984)]. The numbers of independent recombinants in the three libraries were: 2.0×10^6 for the non-CF sweat gland cells, 4.5×10^6 for the CF sweat gland cells, and 3.2×10^6 from T84 cells. Standard procedures were used for screening [T. Maniatis, E. F. Fritsch, J. Sambrook, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982)]. Bluescript plasmids were rescued from plaque-purified clones by excision with M13 helper phage (Stratagene). The lung and pancreas libraries were purchased from Clontech Lab Inc. (Catalog Nos. HL1066b and HL1069h, respectively).
11. The start point of the CF gene transcript was derived by primer extension procedures [F. J. Calzone, R. J. Britten, E. H. Davidson, *Methods Enzymol.* **152**, 611 (1987)]. The oligonucleotide primer [positioned 157 nucleotides (nt) from the 5′ end of the 10-1 clone] was end-labeled with [γ-^{32}P]ATP (Amersham, 5000 Ci/mmole) and T4 polynucleotide kinase, purified by gel filtration, and annealed with ~5 μg of T84 poly(A)$^+$ RNA for 2 hours at 60°C. The extension reaction was performed at 41°C for 1 hour with avian myeloblastosis virus (AMV) reverse transcriptase (Life Sciences, Inc.) and terminated by addition of NaOH to 0.4*M* and EDTA to 20 m*M*, with subsequent neutralization with ammonium acetate (*p*H 4.6). The products were treated with phenol, precipitated with ethanol, redissolved in buffer with formamide, and analyzed on a polyacrylamide sequencing gel.
12. The anchored PCR procedure [M. A. Frohman, M. K. Dush, G. R. Martin, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8998 (1988)] was used to synthesize cDNA corresponding to the 5′ and 3′ ends of the transcript. For the 5′ end clones, poly(A)$^+$ RNA from pancreas and T84 cells were subjected to reverse-transcription with the use of an exon 2–specific primer (*11*). The first strand cDNA products were fractionated on an agarose column and the fractions containing large species were identified by gel electrophoresis after the polymerase chain reaction [R. K. Saiki *et al.*, *Science* **230**, 1350 (1985)] with a pair of oligonucleotide primers (145 nt apart within the 10-1 sequence) just 5′ of the extension primer. These products were pooled, concentrated, and treated with terminal deoxynucleotidyl transferase (BRL) and dATP, as recommended by the supplier. Second strand synthesis was performed with Taq Polymerase (Cetus, AmpliTaq) and an oligonucleotide containing a linker sequence, 5′-CGGAATTCTCGAGATC$(T)_{12}$-3′. This linker, together with another primer (internal to the extension primer) with an Eco RI restriction site at its 5′ end, was then used for PCR. After digestion with Eco RI and Bgl II, products were purified and cloned in Bluescript KS (Stratagene) by standard procedures. All the recovered clones contained inserts of more than 350 nt. The 3′ end clones were generated with the use of similar procedures. PCR amplification was carried out with the linker described above and an oligonucleotide with the sequence 5′-ATGAAGTCCAAGGATTTAG-3′, which is ~70 nt upstream of the Hind III site at position 5027 (Fig. 2). The products were digested with Hind III and Xho I and cloned in the Bluescript vector. Candidate clones were identified by hybridization with the 3′ end of cDNA T16-4.5. All PCR's were performed for 30 cycles as described by the enzyme supplier.
13. M. Kozak, *Nucleic Acids Res.* **12**, 857 (1984); *Cell* **44**, 283 (1986).
14. Other sequence differences were noted between the normal (T16-4.5) and CF (C1-1/5) cDNA clones. At position 2629 (Fig. 2), T16-4.5 contained a C and C1-1/5 a T, resulting in a change of Leu to Phe. At position 4555, the base was G in T16-4.5 but A in C1-1/5 (Val to Met). The differences may be results of cDNA cloning artifacts or may represent sequence polymorphisms. Specific oligonucleotide hybridization analysis of patient or family DNA should distinguish these possibilities. Since these changes are conserved amino acid substitutions, they are unlikely to be causative mutations. Additional nucleotide differences were observed in the 3′ untranslated region between different cDNA clones and the corresponding genomic DNA sequence.
15. C. F. Higgins, M. P. Gallagher, M. L. Mimmack, S. R. Pearce, *BioEssays* **8**, 111 (1988); C. F. Higgins *et al.*, *Nature* **323**, 448 (1986).
16. P. Gros, J. Croop, D. E. Housman, *Cell* **47**, 371 (1986); C. Chen *et al.*, *ibid.*, p. 381; J. H. Gerlach *et al.*, *Nature* **324**, 485 (1986); P. Gros, M. Raymond, J. Bell, D. Housman, *Mol. Cell. Biol.* **8**, 2770 (1988).
17. Several large families of integral membrane proteins are known, including: (i) A number of ligand-gated ion channels of which the nicotinic acetylcholine receptor [R. M. Stroud and J. Finer-Moore, *Annu. Rev. Cell Biol.* **1**, 317 (1985)] is the prototype. Receptors for the inhibitory neurotransmitters GABA (*33*) and glycine are included in this family. (ii) A family of ion channels with a totally different structural motif are the voltage-gated, sodium, calcium, and potassium channels (*27*). (iii) Involved in the translocation of ions are the structurally related cation pumps such as the Ca^{2+}-ATPase [C. J. Brandl, N. M. Green, B. Korczak, D. H. MacLennan, *Cell* **44**, 597 (1986)], the Na^+,K^+-ATPase [G. E. Shull and J. B. Lingrel, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 4039 (1987)], and the H^+,K^+-ATPase [G. E. Shull and J. B. Lingrel, *J. Biol. Chem.* **261**, 16788 (1986)]. These are but examples.
18. G. B. Wilson, T. L. Jahn, J. R. Fonseca, *Clin. Chim. Acta* **49**, 79 (1973); V. van Heyningen, C. Hayward, J. Fletcher, C. McAuley, *Nature* **315**, 513 (1985).
19. J. Garnier, D. J. Osguthorpe, B. Robson, *J. Mol. Biol.* **120**, 97 (1978); J. Kyte and R. F. Doolittle, *ibid.* **157**, 105 (1982).
20. D. D. Pless and W. J. Lennarz, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 134 (1977).
21. P. J. Blackshear, A. C. Nairn, J. F. Kuo, *FASEB J.* **2**, 2957 (1989).
22. J. Hess, W. Wels, M. Vogel, W. Goebel, *FEMS Microbiol. Lett.* **34**, 1 (1986).
23. K. O'Hare, C. Murphy, R. Levis, G. M. Rubin, *J. Mol. Biol.* **180**, 437 (1984).
24. M. J. Friedrich, L. C. DeVeaux, R. J. Kadner, *J. Bacteriol.* **167**, 928 (1986).
25. K. Ohyama *et al.*, *Nature* **322**, 572 (1986).
26. G. F.-L. Ames, *Annu. Rev. Biochem.* **55**, 397 (1986).
27. M. Noda *et al.*, *Nature* **320**, 188 (1986); T. Tanabe *et al.*, *ibid.* **328**, 313 (1987); A. Baumann *et al.*, *EMBO J.* **6**, 3419 (1987).
28. C.-M. Chen, T. K. Misra, S. Silver, B. P. Rosen, *J. Biol. Chem.* **261**, 15030 (1986).
29. J. Krupinski *et al.*, *Science* **244**, 1558 (1989).
30. In addition to the major NBF homologies, searches of the PIR and SWISSPROT data bases detected shorter stretches of sequence homology with other proteins including the following:

```
CFTR 122-136     Y L G I G L C L L F I V R T L
                 : :   :   :   :   : : . :   :
GLNP 198-212     Y L I I T L V L S F I L R R L

CFTR 307-319     S S A F F F S G F F V V F
                 :     : : : . : : : .   :
COX   89-101     S E V F F F A G F F W A F

CFTR 701-713     I L N P I N S I R K F S I
                 : :   : . :   : : : . . :
NaCh 111-123     I L T P F N P I R K L A I

CFTR 1425-1442   D S I Q K L L N E R S L F R Q A I S
                 : : :   : :     : :   : :   :   . :
raf   578-595    D S I K K L R D E R P L F P Q I L S
```

GLNP, glutamine permease of *E. coli* [T. Nohno, T. Saito, J. Hóng, *Mol. Gen. Genet.* 205, 260 (1986)]; COX, human cytochrome c oxidase polypeptide III [S. Anderson *et al.*, *Nature* 290, 457 (1981)]; NaCh, rat brain sodium channel III (*32*); raf, the serine-threonine kinase proto-oncogene of *Xenopus laevis* (*31*). The first two sequences are within membrane spanning segments and probably reflect only coincidental arrangements of the hydrophobic residues suited to this function. In contrast, the latter two sequences are both in polar hydrophilic regions of the proteins. The large extent of amino acid conservation (11 of 13 residues) implies some functional relation between these short segments of the primary structure of the Na^+ channel and CFTR. Similarities between sequences at the same relative locations with respect to the COOH-termini of the raf kinase and CFTR suggest that they may also share at least a small facet of their structures and functions.

31. R. LeGuellec, K. LeGuellec, J. Paris, M. Philippe, *Nucleic Acids Res.* 16, 10357 (1988).
32. M. Noda *et al.*, *Nature* 312, 121 (1984); L. Salkoff *et al.*, *Science* 237, 744 (1987).
33. P. R. Schofield *et al.*, *Nature* 328, 221 (1987).
34. F. Sanger, S. Nicklen, A. R. Coulsen, *Proc. Natl. Acad. Sci. U.S.A.* 74, 5463 (1977).
35. D. Eisenberg, E. Schwarz, M. Komaramy, R. Wall, *J. Mol. Biol.* 179, 125 (1984).
36. A. P. Feinberg and B. Vogelstein, *Anal. Biochem.* 132, 6 (1983).
37. J. Rommens *et al.*, *Am. J. Hum. Genet.* 43, 645 (1988).
38. S. J. Foote *et al.*, *Cell* 57, 921 (1989).
39. J. P. McGrath and A. Varshavsky, *Nature* 340, 400 (1989).
40. B. P. Surin, H. Rosenberg, G. B. Cox, *J. Bacteriol.* 161, 189 (1985).
41. C. F. Higgins *et al.*, *Nature* 298, 723 (1982).
42. E. Gilson, H. Nikaido, M. Hofnung, *Nucleic Acids Res.* 10, 7449 (1982).
43. I. D. Hiles, M. P. Gallagher, D. H. Jamieson, C. F. Higgins, *J. Mol. Biol.* 195, 125 (1987).
44. A. W. Bell *et al.*, *J. Biol. Chem.* 261, 7652 (1986).
45. R. F. Doolittle *et al.*, *Nature* 323, 451 (1986).
46. I. J. Evans and J. A. Downie, *Gene* 43, 95 (1986).
47. D. R. Gill, G. H. Hatfull, G. P. C. Salmond, *Mol. Gen. Genet.* 205, 134 (1986).
48. J. E. Walker, M. Saraste, M. J. Runswick, N. J. Gay, *EMBO J.* 1, 945 (1982).
49. D. C. Fry, S. A. Kuby, A. S. Mildvan, *Proc. Natl. Acad. Sci. U.S.A.* 83, 907 (1986).

50. We thank O. Augustinas for the collection of tissues; T. Jensen and R. Baird for the culturing of epithelial cells; L. Naismith for the isolation of RNA; D. Kennedy and D. Markiewicz for technical assistance and M. Buchwald and M. Dean for discussions. Supported by NIH grants DK39690 (F.S.C.) and DK34944 (L.C.T.), the Cystic Fibrosis Foundation (U.S.A.), the Canadian Cystic Fibrosis Foundation, and the Sellers Fund. J.M.R. holds a postdoctoral fellowship from the Medical Research Council (MRC) of Canada and F.S.C. is an Associate Investigator of the Howard Hughes Medical Institute.

7 August 1989; accepted 18 August 1989

Identification of the Cystic Fibrosis Gene: Genetic Analysis

BAT-SHEVA KEREM, JOHANNA M. ROMMENS, JANET A. BUCHANAN, DANUTA MARKIEWICZ, TARA K. COX, ARAVINDA CHAKRAVARTI, MANUEL BUCHWALD, LAP-CHEE TSUI

Approximately 70 percent of the mutations in cystic fibrosis patients correspond to a specific deletion of three base pairs, which results in the loss of a phenylalanine residue at amino acid position 508 of the putative product of the cystic fibrosis gene. Extended haplotype data based on DNA markers closely linked to the putative disease gene locus suggest that the remainder of the cystic fibrosis mutant gene pool consists of multiple, different mutations. A small set of these latter mutant alleles (about 8 percent) may confer residual pancreatic exocrine function in a subgroup of patients who are pancreatic sufficient. The ability to detect mutations in the cystic fibrosis gene at the DNA level has important implications for genetic diagnosis.

ALTHOUGH THE FREQUENCY OF CYSTIC FIBROSIS (CF) IS not uniformly high among all Caucasian populations, a consensus estimate is that it occurs once in 2000 live births (*1*). On the basis of the autosomal recessive mode of inheritance for this disease, a mutant allele frequency of 0.022 may be derived. Several different mechanisms, including high mutation rate (*2*), heterozygote advantage (*3*), genetic drift (*4*), multiple loci (*5*), and reproductive compensation (*6*), have been proposed in attempts to explain the high incidence and, indirectly, the nature of the CF mutations. Although some of these hypotheses could not be further addressed because of the lack of knowledge about the basic defect in CF, several important observations have been made during the past few years through genetic analysis of the families of affected individuals (*7–20*).

Extensive linkage analysis provides evidence for the existence of a single CF locus on human chromosome 7 (region q31) (*7–10, 21*). The detection of allelic and haplotype association between the CF

B. Kerem, J. M. Rommens, J. A. Buchanan, D. Markiewicz, M. Buchwald and L.-C. Tsui are in the Department of Genetics, Research Institute, The Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. T. K. Cox and A. Chakravarti are in the Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA 15261. M. Buchwald and L.-C. Tsui are also members of the Departments of Medical Genetics and Medical Biophysics, University of Toronto, Toronto, Ontario M5S 1A8, Canada.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants:	Gregory et. al.	Examiner:	Dr. K. C. Carlson
Serial No.:	08/087,132	Art Unit:	1812
Filed:	July 2, 1993	Docket:	NZI-012CN IG4-9.2(FWC)

For: NEW DIAGNOSTIC AND TREATMENT METHODS INVOLVING THE CYSTIC FIBROSIS TRANSMEMBRANE REGULATOR

Honorable Commissioner of
Patents and Trademarks
Washington, DC 20231

EXHIBIT BOOK — PART 1 of 2

S I R:

This Exhibit Book is submitted in connection with Applicants' Reply with Amendment of June 9, 1995 to the Offical Action that is pending herein

The Documents bound herein are:

(A) J. M. Rommens et al., "Identification of the Cystic Fibrosis Gene: Chromosome Walking and Jumping", Science, 245, 1989, pp. 1059-1065.

(B) J. Riordan et al., "Identification of the Cystic Fibrosis Gene: Cloning and Characterization of Complementary DNA", Science, 245, 1989, pp. 1066-1073.

(C) L. Tsui et al., "Cystic Fibrosis Gene", International Patent Application Number PCT/CA90/00267 published on March 7, 1991, bearing Publication Number WO 91/002796, and claiming the priority of United States patent applications 07/396,894, 07/399,945, and 07/401,609 filed, respectively, on August 22, 24, and 31, 1989.

(D) L. Tsui. et al. "Introns and Exons of the Cystic Fibrosis Gene and Mutations at Various Positions of the Gene", International Patent Application Number PCT/CA91/00009 published on July 25, 1991, bearing Publication Number WO 91/10734, and claiming the priority of Canadian national applications 2,007,699, 2,011,253, and 2,020,817 filed, respectively, on January 12, March 1, and July 10, 1990.

(E) F.S. Collins et al.,"Gene Therapy for Cystic Fibrosis", International Patent Application Number PCT/US91/06660 published on April 2, 1992, bearing Publication Number WO 92/05273, and claiming the priority of United States patent application 07/584,275 filed on September 18, 1990.

(F) United States Patent 5,240,846 to Collins et al. issued August 31, 1993 from Application No. 07/584,275 filed September 18, 1990

(G) L. Tsui et al., "Stable Propagation of Modified Full Length Cystic Fibrosis Transmembrane Conductance Regulator Protein cDNA in Heterologous Systems", International Patent Application Number PCT/CA91/00341 published on April 2, 1992, bearing Publication Number WO 92/05252, and claiming the priority of Great Britain national application 9020632.7 filed September 21, 1990.

(H) M. Drumm et al., "The Full Length CFTR cDNA is Toxic in Bacteria", Pediatric Pulmonology , Supplement 5 (Abstracts), October 1990, Abstract No. 8, page 189.

(I) M.L. Drumm et al., "Correction of the Cystic Fibrosis Defect in vitro by Retrovirus-Mediated Gene Transfer", Cell, 62, September 21, 1990, pp. 1227-1233.

(J) L. Tsui. "Probing the Basic Defect in Cystic Fibrosis", Current Opinion in Genetics and Development , 1, 1991, pp. 4-10.

(K) Ex parte Gray, Board of Patent Appeals and Interferences, 10 USPQ 2d 1929, 1989.

(L) Ex parte Forman, Board of Patent Appeals and Interferences, 230 USPQ 546, 1986.

(M) In re Wands, Court of Appeals, Federal Circuit, 8 USPQ 2d 1400, 1988.

(N) Ex parte Mark, Board of Patent Appeals and Interferences, 12 USPQ 2d 1904, 1989.

(O) Atlas Powder Co. v. E.I. DuPont Nemours, 750 F.2d 1569, 224 USPQ 409 (Fed. Cir. 1984).

Respectfully submitted,
GENZYME CORPORATION

Dated: 6/9/95

By: William G. Gosz
William G. Gosz, Esquire
Registration No. 27,787
Attorney for Applicants

One Mountain Road
Framingham, MA 01701- 9322
(508) 872-8400

Identification of the Cystic Fibrosis Gene: Chromosome Walking and Jumping

JOHANNA M. ROMMENS, MICHAEL C. IANNUZZI, BAT-SHEVA KEREM, MITCHELL L. DRUMM, GEORG MELMER, MICHAEL DEAN, RICHARD ROZMAHEL, JEFFERY L. COLE, DARA KENNEDY, NORIKO HIDAKA, MARTHA ZSIGA, MANUEL BUCHWALD, JOHN R. RIORDAN, LAP-CHEE TSUI, FRANCIS S. COLLINS

An understanding of the basic defect in the inherited disorder cystic fibrosis requires cloning of the cystic fibrosis gene and definition of its protein product. In the absence of direct functional information, chromosomal map position is a guide for locating the gene. Chromosome walking and jumping and complementary DNA hybridization were used to isolate DNA sequences, encompassing more than 500,000 base pairs, from the cystic fibrosis region on the long arm of human chromosome 7. Several transcribed sequences and conserved segments were identified in this cloned region. One of these corresponds to the cystic fibrosis gene and spans approximately 250,000 base pairs of genomic DNA.

CYSTIC FIBROSIS (CF) IS REGARDED AS THE MOST COMMON severe autosomal recessive disorder in the Caucasian population, with a disease frequency of 1 in 2000 live births and a calculated carrier frequency of about 5 percent (*1*). The major clinical symptoms and signs include chronic pulmonary disease, pancreatic exocrine insufficiency, and an increase in the concentration of sweat electrolytes. Although recent advances have been made in the analysis of ion transport across the apical membrane in CF epithelium (*2*), it is not clear that the abnormal regulation of chloride channels represents the primary lesion in the disease. Apart from these electrophysiological studies, an alternative approach has been taken in an attempt to understand the nature of the molecular defect through direct cloning of the responsible gene on the basis of its chromosomal location (*3, 4*).

Linkage analysis based on a large number of polymorphic DNA markers has unambiguously assigned the CF locus (*CF*) to the long arm of chromosome 7, band q31 (*3–5*). The identification of closely linked flanking markers, *MET* and *D7S8*, has made it possible to use various novel gene cloning strategies to pinpoint the CF gene. These methods include chromosome jumping from the flanking markers (*6*), cloning of DNA fragments from a defined physical region with the use of pulsed field gel electrophoresis (*7*), a combination of somatic cell hybrid and molecular cloning techniques designed to isolate DNA fragments from undermethylated CpG islands near *CF* (*8*), chromosome microdissection and cloning (*9*), and saturation cloning of a large number of DNA markers from the 7q31 region (*10*).

The saturation mapping approach, by systematic examination of DNA markers from a flow-sorted genomic DNA library specific to chromosome 7, allowed the identification of two additional DNA markers (*D7S122* and *D7S340*) closely linked to *CF* (*10*). Genetic and physical mapping studies indicated the order of the four markers to be *MET-D7S340-D7S122-D7S8*, with distance intervals of 500, 10, and 980 kilobase (kb) pairs, respectively (*11*). This distance estimate for the *MET-D7S8* interval agrees well with the data from previous genetic (*4, 5, 10*) and physical mapping (*12*) studies.

Chromosome walking and jumping. As the genetic data indicated that *D7S122* and *D7S340* were probably in close proximity to *CF*, and the physical map of the region was well defined, the next logical step was to clone a large amount of the surrounding DNA and search for candidate gene sequences. In addition to conventional chromosome walking methods, the chromosome jumping technique was used to accelerate the process, as a new bidirectional walk could be initiated from the end point of each jump. Furthermore, sequential walks halted by "unclonable" regions often encountered in the mammalian genome could be circumvented by chromosome jumping (see below). Parallel chromosome jumping experiments were also performed from *D7S8* toward *D7S122* and *D7S340* to narrow the region of interest (*13*).

Ten genomic libraries were constructed during the course of our experiments (*14*). The contiguous chromosome region covered by chromosome walking and jumping was about 280 kb (Fig. 1). This effort involved the isolation and characterization of 49 recombinant phage and cosmid clones and nine jumping clones. The ability to bias the direction of jumps by careful choice of probes (*6*) proved to be a useful feature of the strategy.

A restriction map of the cloned human DNA segments derived from chromosome walking and jumping was constructed (Fig. 1). As the two independently isolated DNA markers, *D7S122* (pH131)

J. M. Rommens, B. Kerem, G. Melmer, R. Rozmahel, D. Kennedy, M. Zsiga, M. Buchwald, and L.-C. Tsui are in the Department of Genetics, Research Institute, The Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. M. C. Iannuzzi, M. L. Drumm, J. L. Cole, N. Hidaka, and F. S. Collins are at the Howard Hughes Medical Institute and Departments of Internal Medicine and Human Genetics, University of Michigan, Ann Arbor, MI 48109. M. Buchwald and L.-C. Tsui are also faculty members of the Departments of Medical Genetics and Medical Biophysics, University of Toronto, Toronto M5S 1A8, Canada. M. Dean is at Program Resources, Inc., Cancer Research Facility, Frederick, MD 21701. J. R. Riordan is in the Department of Biochemistry, Research Institute, The Hospital for Sick Children, and the Departments of Biochemistry and Clinical Biochemistry, University of Toronto M5S 1A8, Canada.

and *D7S340* (TM58), were only ~10 kb apart (Fig. 1), the walks and jumps were essentially initiated from a single point. The direction of walking and jumping with respect to *MET* and *D7S8* was then established with the crossing of several rare-cutting restriction endonuclease recognition sites, such as those for Xho I, Nru I, and Not I (Fig. 1), and with reference to the long-range physical map (*11, 12*). The pulsed field mapping data also revealed that the Not I site identified in our study (see Fig. 1, position 113 kb) corresponded to the one previously found associated with the *int*-related protein (IRP) locus (*IRP*) (*8*). As subsequent genetic studies showed that *CF* was most likely located between *IRP* and *D7S8* (*15, 16*), our walking and jumping effort, as described below, was directed exclusively toward cloning of this interval.

Three regions in the 280-kb segment were not readily recoverable in the amplified genomic libraries initially used (*14*). These less clonable regions were located near the DNA segments H2.3A and X.6 and just beyond cosmid cW44, at positions 75 to 100 kb, 205 to 225 kb, and 275 to 285 kb, respectively (Fig. 1). The recombinant clones near H2.3A were unstable, and underwent dramatic rearrangements after only a few passages of bacterial culture. To fill in the resulting gaps, we constructed primary walking libraries with special host-vector systems that allow propagation of unstable sequences (*14, 17*). Although the region near cosmid cW44 has not yet been recovered, the region near X.6 was successfully rescued with these libraries. Mammalian DNA segments with unusual secondary structure or repetitive elements are unstable in bacterial cells (*17*), but the nature of the less clonable sequences encountered in our study remain to be determined. It is of interest that potential recombination hot spots have been identified near H2.3A and the end of cW44 (*16*).

Alignment of cloned regions with genomic DNA. Together with the genomic DNA sequences isolated with the overlapping cDNA clones described by Riordan *et al.* (*18*), the entire region cloned in our study extended >500 kb. To ensure that the DNA segments isolated by the chromosome walking and jumping procedures were colinear with the genomic sequence, we examined each segment by (i) hybridization analysis with human-rodent somatic hybrid cell lines to confirm localization on chromosome 7 (*10, 19*); (ii) pulsed field gel electrophoresis; and (iii) comparison of the restriction map of the cloned DNA to that of the genomic DNA. Accordingly, single copy human DNA sequences were isolated from each recombinant phage and cosmid clone and were used as probes in each of these hybridization analyses (*20, 21*).

Although most phage and cosmid isolates represented correct walk and jump clones, a few resulted from cloning artifacts or cross-hybridizing sequences from other regions in the human genome, or from the hamster genome in cases where the libraries were derived from a human-hamster hybrid cell line. Confirmation of correct localization was particularly important for clones isolated by chromosome jumping. Because this cloning strategy requires the ligation of the two ends of a large genomic segment (*6*), tandem ligations of unrelated molecules can give rise to anomalous jumping clones. One of the jump clones was not located on chromosome 7 and was discarded.

Further confirmation of the overall physical map of the overlapping clones was obtained by long-range restriction mapping with the use of pulsed field gel electrophoresis (*11, 12*). A preliminary long-range map of this region describing *D7S122* and *D7S340* was previously published (*11*). The more recent walk-jump clones and cDNA clones corresponding to the CF locus generated a more extensive pulsed field restriction map, which was in complete concordance with that derived from chromosome walking (Fig. 2). Many of the recognition sites for rare-cutting restriction enzymes in this region, such as Not I and Bss HII, were resistant to digestion in the human-rodent cell hybrid (*19*), presumably due to DNA methylation. These sites were less resistant to digestion, however, in other human cell lines (Fig. 2).

The result of the long-range restriction mapping study showed that the entire CF locus was contained on a 380-kb Sal I fragment (Fig. 2). Alignment of the restriction sites derived from pulsed field gel analysis to those identified in the partially overlapping genomic DNA clones revealed that the size of the CF locus was about ~250 kb.

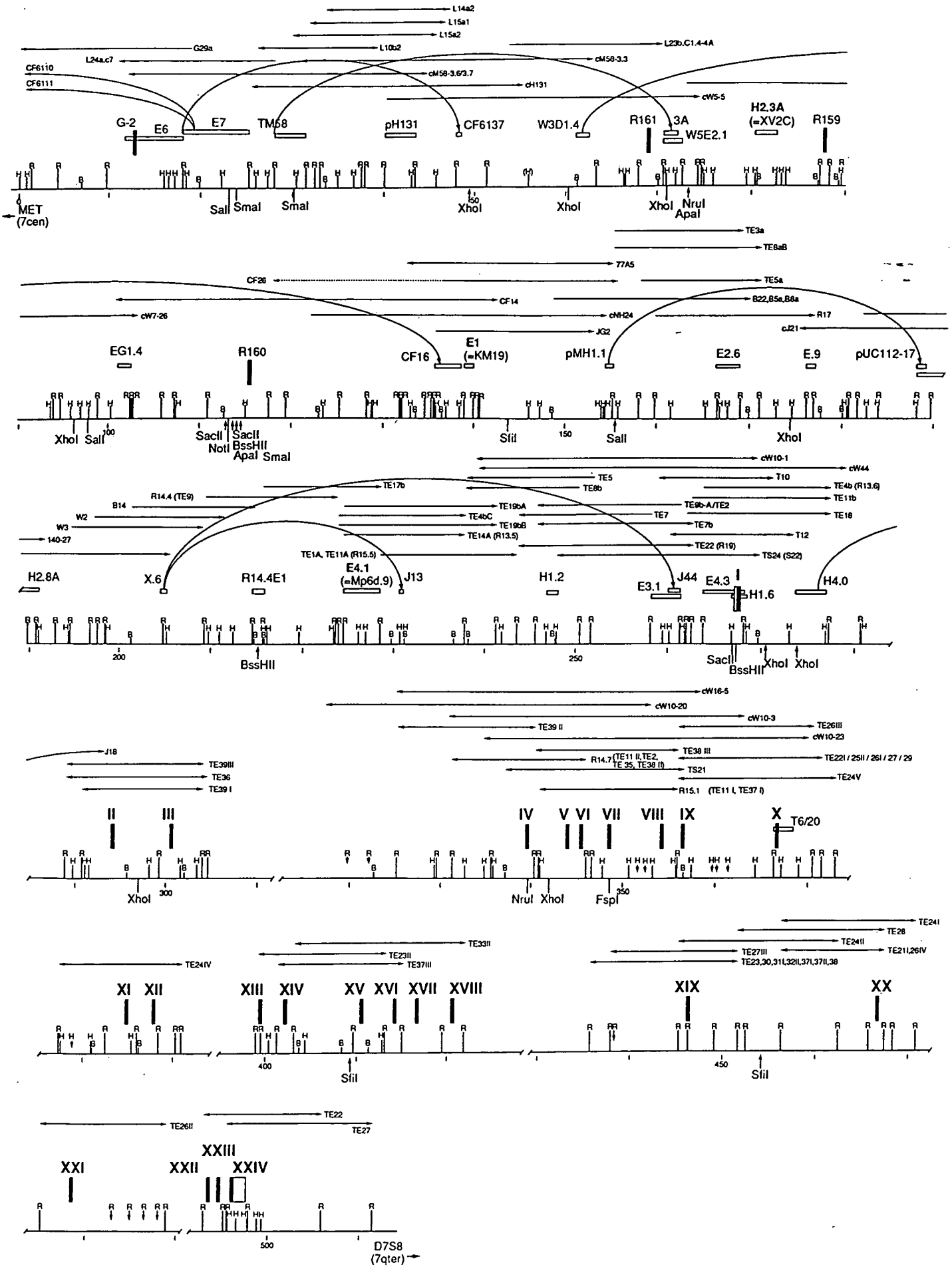
The most informative restriction enzyme that served to align the map of the cloned DNA fragments and the long-range restriction map was Xho I; all of the nine Xho I sites identified with the recombinant DNA clones appeared to be susceptible to at least partial cleavage in genomic DNA (compare maps in Figs. 1 and 2). Furthermore, hybridization analysis with probes derived from the 3′ end of the CF locus identified two Sfi I sites and confirmed the position of an anticipated Nae I site.

Search for gene sequences. A positive result based on one or more of the following criteria suggested that a cloned DNA segment might contain candidate gene sequences: (i) detection of cross-hybridizing sequences in other species (as many genes show evolutionary conservation) (*22*), (ii) identification of CpG islands, which often mark the 5′ end of vertebrate genes (*23*), (iii) examination of possible mRNA transcripts in tissues affected in CF patients, (iv) isolation of corresponding cDNA sequences, and (v) identification of open reading frames by direct sequencing of cloned DNA segments. The ongoing genetic analysis (*15, 16, 24*) strongly influenced how extensively a region was examined for possible gene sequences. All the methods have potential inherent limitations; because the DNA hybridization method for detecting conserved DNA sequences across species was relatively straightforward and has been successful in detection of other loci (*22*), it was generally used as the first step.

In some of the cross-species hybridization experiments, it was possible to use entire phage or cosmid clones containing human sequences as probes without removal of the repetitive elements because these sequences are in general not shared between distantly related species. Distinct cross-hybridization signals were detected with probes from four regions in the 280-kb span (Fig. 3).

Conserved region 1 was defined by the DNA segment G-2 (position 13 in Fig. 1); region 2 was detected by the cosmid CF14 (positions 100 to 142); region 3 was defined by the probe R14.4E1

Fig. 1 (facing page). Restriction map of the region of chromosome 7 containing *CF*. The map proceeds from left to right in six tiers with the direction of ends toward 7cen and 7qter as indicated. The restriction map for the enzymes Eco RI (R), Hind III (H), and Bam HI (B) is shown above the solid line, spanning the entire cloned region. Restriction sites indicated with arrows rather than vertical lines indicate sites that have not been unequivocally positioned. Additional restriction sites for other enzymes are shown below the line. The scale is in kilobases. Gaps in the cloned region are indicated by a gap in the solid line (//). These occur only in the portion detected by cDNA clones of the CF transcript and, on the basis of pulsed field mapping of the region (Fig. 2), are unlikely to be large. Chromosome jumps are indicated by the arcs. Walking clones are indicated by horizontal arrows above the map, with the direction of the arrow indicating the walking progress obtained with each clone. Cosmid clones begin with C or c; all other clones are phage. Cosmid CF26 proved to be a chimera; the dashed portion is derived from a different genomic fragment on another chromosome. Roman numerals I through XXIV indicate the location of exons of the CF gene. The horizontal boxes shown above the line are probes used in this and accompanying papers (*16, 18*). Three of the probes represent independent subcloning of fragments previously identified to detect polymorphisms in this region: H2.3A corresponds to probe XV2C (*8*), probe E1 corresponds to KM19 (*8*), and E4.1 corresponds to Mp6d.9 (*37*). G-2 is a subfragment of E6 that detects a transcribed sequence (see Fig. 4); R161, R159, and R160 are synthetic oligonucleotides constructed from the IRP locus sequence (*26*), indicating the location of this transcript on the genomic map.


L14a2
L15a1
L15a2
G29a
L10b2
L23b,C1.4-4A
L24a,c7
cM58-3.3
CF6110
cM58-3.6/3.7
CF6111
cH131
cW5-5
H2.3A
(=XV2C)
G-2
E6
E7
TM58
pH131
CF6137
W3D1.4
R161
3A
W5E2.1
R159
MET
(7cen)
SalI
SmaI
SmaI
XhoI
XhoI
XhoI
NruI
ApaI
TE3a
TE8aB
77A5
CF26
TE5a
CF14
B22,B5a,B8a
cW7-26
cNH24
R17
JG2
cJ21
EG1.4
R160
CF16
E1
(=KM19)
pMH1.1
E2.6
E.9
pUC112-17
XhoI
SalI
SacII
NotI
SacII
BssHII
ApaI
SmaI
SfiI
SalI
XhoI
cW10-1
cW44
TE5
T10
TE17b
TE8b
TE4b (R13.6)
R14.4 (TE9)
TE11b
B14
TE19bA
TE9b-A/TE2
W2
TE4bC
TE7
TE18
W3
TE19bB
TE7b
140-27
TE14A (R13.5)
T12
TE1A, TE11A (R15.5)
TE22 (R19)
TS24 (S22)
H2.8A
X.6
R14.4E1
E4.1
(=Mp6d.9)
J13
H1.2
E3.1
J44
E4.3
H1.6
H4.0
BssHII
SacII
BssHII
XhoI
XhoI
cW16-5
cW10-20
cW10-3
TE39 II
TE26III
cW10-23
J18
TE38 III
R14.7 (TE11 II,TE2, TE 35, TE38 II)
TE22I / 25II / 26I / 27 / 29
TE39III
TS21
TE36
TE24V
TE39 I
R15.1 (TE11 I, TE37 I)
II
III
IV
V
VI
VII
VIII
IX
X
T6/20
XhoI
NruI
XhoI
FspI
TE24I
TE28
TE33II
TE24II
TE23II
TE27III
TE21I,26IV
TE24IV
TE37III
TE23,30,31I,32II,37I,37II,38
XI
XII
XIII
XIV
XV
XVI
XVII
XVIII
XIX
XX
SfiI
SfiI
TE22
TE26II
TE27
XXIII
XXI
XXII
XXIV
D7S8
(7qter)

(position 215); and region 4 was initially recognized by the probes E4.3 and H1.6 (positions 264 to 268). The DNA segments that revealed sequence conservation were then tested for RNA hybridization and used to screen cDNA libraries of tissues affected in CF. Only a brief description of regions 1 to 3 is given below; region 4 corresponds to the 5′ end of the CF locus.

The probe G-2, one of the first segments tested, detected a 3.7-kb transcript in simian virus 40 (SV40)–transformed human fibroblasts (Fig. 4A). When this fragment was used to screen a human fibroblast and a human lung cDNA library, three independent clones were isolated (*25*). The overlapping cDNA clones spanned a length of 1.8 kb, and nucleotide sequence analysis revealed a potential open reading frame corresponding to the 3′ end of a coding region. Alignment of the cDNA sequence with that of the genomic DNA showed perfect sequence identity as well as exon-intron structures. Because this gene could not be the CF gene on the basis of genetic data (*16*), characterization studies were not continued.

Region 2 was identified by the cosmid clone CF14, which revealed strong cross-species hybridization signals in mouse, chicken, and bovine DNA (Fig. 3A). Restriction mapping of the genomic DNA showed that part of this region corresponded to the previously reported *IRP* (*8, 26*). The extent of this locus was subsequently confirmed by hybridization with oligonucleotide probes made to the IRP sequence of Wainwright *et al.* (*26*) (Fig. 1). As family studies indicated that *CF* maps to the D7S8 side of *IRP* (*15, 16*), chromosome walking and jumping experiments were continued in this direction.

The first region that revealed a transcript at a location on the D7S8 side of the IRP gene was identified by the probe CF16 (Fig. 1, positions 135 to 140). This probe detected RNA transcripts of different sizes in various tissues; a 2-kb species was observed in tracheal epithelium and pancreas, a less abundant 4-kb mRNA was seen in the brain, and a 9-kb transcript was observed in the liver (Fig. 4). When this probe was used to screen cDNA libraries made from human lung and cultured epithelial cells from sweat glands, more than ten clones were isolated. Restriction enzyme analyses of a subset of these cDNA clones revealed significant differences. Nucleotide sequence analyses of representative cDNA clones and the genomic DNA revealed that they shared a high degree (more than 85 percent) of sequence similarity but that none of the cDNA clones showed perfect identity with the genomic DNA sequence. Furthermore, neither the genomic DNA nor any of the cDNA clones contained an open reading frame. Screening a sequence databank (GenBank) showed that these clones share remarkable sequence similarity with a region in the β-globin locus (between ε and $^{G}\gamma$), suggesting that these sequences correspond to a transcribed repetitive DNA family that is distinct from the LINE-1 (long interspersed element) sequence (*27*).

Region 3 (position 215) contained a high proportion of CpG nucleotide residues, as determined by sequencing the 1-kb Eco RI fragment of genomic DNA. Open reading frames were also detected; however, neither RNA transcripts nor cDNA clones were detected with this probe. This could indicate that this transcript is restricted in tissue or developmental specificity, or that the notably weaker hybridization signals observed in other mammalian DNA's

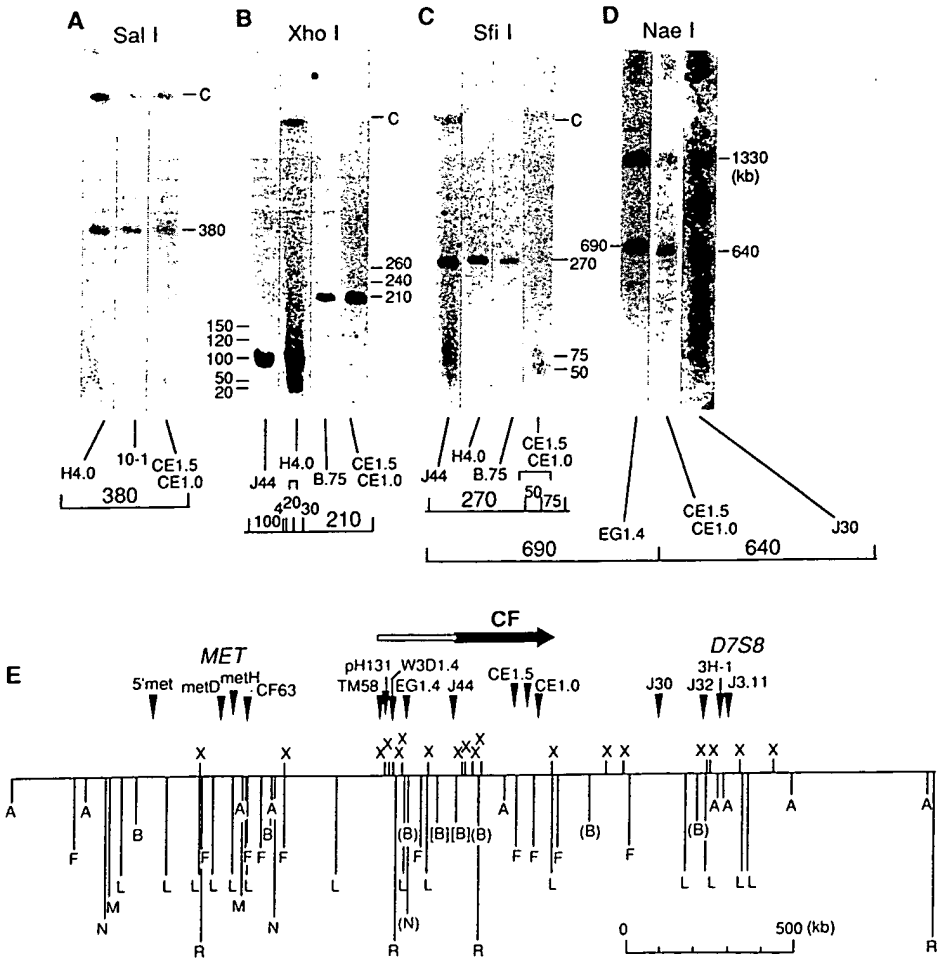
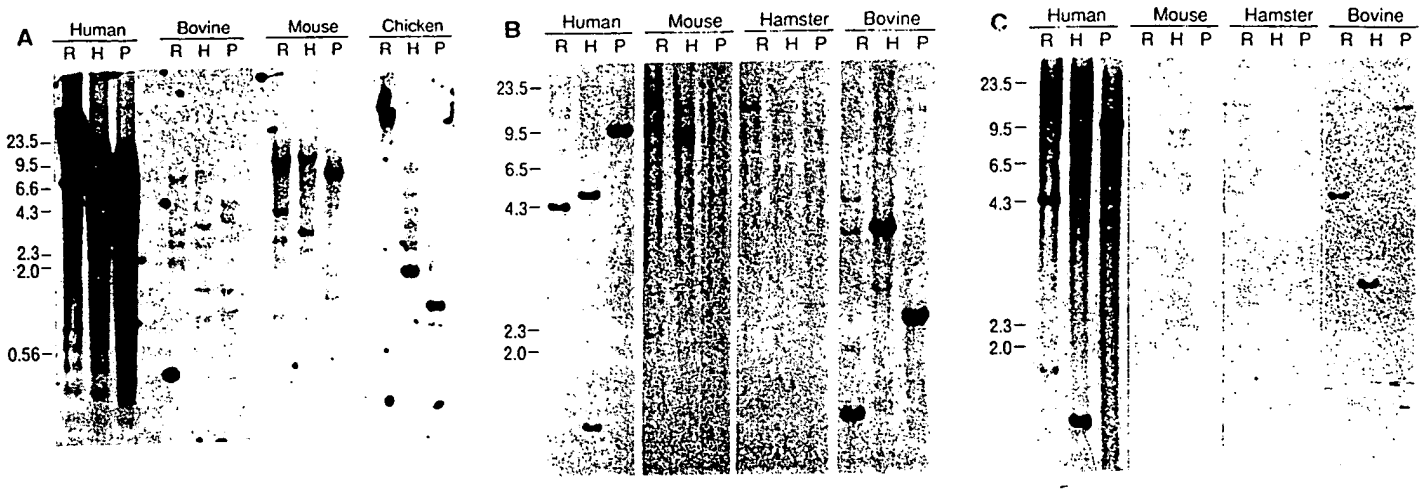


Fig. 2. Pulsed field gel electrophoresis mapping of the cloned region. DNA from the human-hamster cell line 4AF/102/K015 was digested with the enzymes (**A**) Sal I, (**B**) Xho I, (**C**) Sfi I, and (**D**) Nae I, and the fragments were separated by pulsed field gel electrophoresis and transferred to Zetaprobe (Bio-Rad). For each enzyme, a single blot was sequentially hybridized with the probes indicated below each panel, with stripping of the blot between hybridizations. DNA preparation, restriction enzyme digestion, and crossed field gel electrophoresis methods were as described (*11*). Electrophoresis was as follows: in 0.5× TBE (tris, borate, and EDTA) at 7 V/cm for 20 hours with switching linearly ramped from 10 to 40 for (A), (B), and (C), and at 8 V/cm for 20 hours with switching ramped linearly from 50 to 150 for (D). C corresponds to the compression zone region of the gel. Schematic interpretations of the hybridization pattern are given below each panel. Fragment lengths are in kilobases and were sized by comparison to oligomerized bacteriophage λ DNA and *Saccharomyces cerevisiae* chromosomes. Alignment of individual enzyme maps was facilitated by reference to previously described maps (*6, 11*). H4.0, J44, and EG1.4 are genomic probes generated from the walking and jumping experiments (see Fig. 1). J30 was isolated by four consecutive jumps from *D7S8* (*6, 13*). 10-1, B.75, and CE1.5 and CE1.0 together are cDNA probes that cover different regions of the CF transcript: 10-1 contains exons I to VI, B.75 contains exons V to XII, and CE1.5 and CE1.0 together contain exons XII to XXIV. Shown in (**E**) is a composite map of the entire *MET-D7S8* interval. The open box indicates the segment cloned by walking and jumping, and the arrow indicates the region covered by the CF transcript. The CpG-rich region associated with *D7S23* (*8*) is at the Not I site shown in parentheses. This and other sites shown in parentheses or square brackets are not cut in 4AF/102/K015 but have been observed in human lymphoblast cell lines. The symbols for each enzyme are: A, Nae I; B, Bss HII; F, Sfi I; L, Sal I; M, Mlu I; N, Not I; R, Nru I; and X, Xho I.



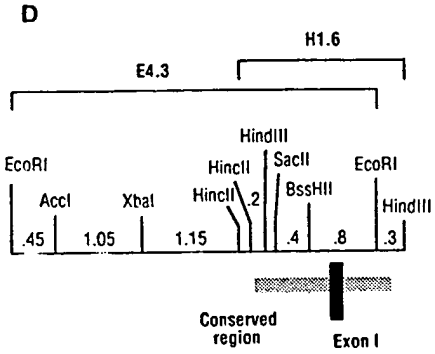


Fig. 3. Detection of conserved sequences by cross-species hybridization. Human, bovine, mouse, hamster, and chicken genomic DNA's were digested with Eco RI (R), Hind III (H), and Pst I (P), and the fragments were subjected to electrophoresis and blotted to Zetabind (Bio-Rad) as described (*10*). The hybridization procedures were also as described (*10*) with the most stringent washing being at 55°C in 0.2× SSC (standard saline citrate) and 0.1 percent SDS. Probes for hybridization (Fig. 1) included: (**A**) entire cosmid CF14, (**B**) E4.3, and (**C**) H1.6. The schematic in (**D**) shows a detailed restriction map of the overlapping segments E4.3 and H1.6. The shaded region indicates the area of cross-species conservation. Sizes are in kilobases.

were due to nonspecific hybridization of the CpG-rich sequence.

The CF locus. The next region of interest was first noted by the strong sequence conservation between human and bovine DNA with the probes E4.3 and H1.6 (Fig. 3, B and C); only weak hybridization was detected in the mouse and hamster DNA with the human probe. The fact that different subsets of bands were detected in bovine DNA with these two overlapping DNA segments suggested that the conserved sequences were located at the boundaries of the overlapped region (Fig. 3D). When these DNA segments were used to detect RNA transcripts from various tissues, no hybridization signal was detected. In an attempt to understand the cross-hybridizing region and to identify possible open reading frames, we determined the DNA sequences of the entire H1.6 and part of the E4.3 fragment. The results showed that, except for a long stretch of CG-rich sequence containing the recognition sites for two restriction enzymes (Bss HII and Sac II) often found associated with undermethylated CpG islands (*23*), there were only short open reading frames, which could not easily explain the strong cross-species hybridization signals.

Undermethylated CpG islands have been associated with the 5′ ends of most housekeeping genes and a number of tissue-specific genes (*23*). To examine the methylation status of the highly CpG-rich region revealed by sequencing, genomic DNA samples prepared from fibroblasts and lymphoblasts were digested with the restriction enzymes Hpa II and Msp I and analyzed by gel-blot hybridization. (The enzyme Hpa II cuts the DNA sequence 5′-CCGG-3′ only when the second cytosine is unmethylated, whereas Msp I cuts this sequence regardless of the state of methylation.) Small DNA fragments were generated by both enzymes, indicating that this CpG-rich region is indeed undermethylated in genomic DNA.

Exhaustive screening of multiple cDNA libraries with the DNA segment H1.6 eventually yielded a single isolate (clone 10-1) carrying a 920-bp insert from a cDNA library constructed from cultured sweat gland cells of a non-CF individual (*18*). Nucleotide sequence analysis indicated that only 113 bp at the 5′ end of this clone aligned with sequences in H1.6 and thus provided a partial explanation for the poor hybridization signals observed in cDNA library screening. Use of the 10-1 cDNA as a probe revealed a 6.5-kb transcript in RNA from the T84 colon cancer cell line (Fig. 4). Results of further cDNA cloning experiments, sequencing, and genetic analysis suggested that H1.6 corresponds to the 5′ end of the gene most likely to be responsible for cystic fibrosis (*16, 18*). With the use of several additional overlapping cDNA clones, a number of genomic DNA segments were isolated from the recombinant phage and cosmid libraries. Alignment of these cloned genomic DNA segments with the long-range restriction map revealed that the locus spans ~250 kb (see Fig. 1). DNA sequencing and gel-blot hybridization demonstrated that this gene locus contains a minimum of 24 exons. Pulsed field analysis of this region from CF patients with a variety of haplotypes (*16*) gave no evidence for any visible genomic rearrangements within this interval (*28*). A detailed description of the coding region of this gene is now available (*18*).

Lessons from the search. A detailed analysis of 280 kb of contiguous DNA isolated by chromosome jumping and walking has permitted the cloning of the locus responsible for cystic fibrosis without prior knowledge of the basic defect. A major difficulty in identifying the CF locus has been the lack of chromosome rearrangements or deletions, which greatly facilitated all previous successes in the cloning of human disease genes by knowledge of map position (*29*). The strategy used in our study may therefore serve as an example for other similar disease gene cloning studies for which no gross genetic alteration has been demonstrated, although the task will be more challenging for disorders that are rare or where diagnosis is difficult.

As discussed above, the use of various molecular cloning techniques has led to the identification of DNA markers closely linked to *CF* (*6–10*). The positioning of these markers relative to each other was facilitated by somatic cell hybrid mapping (*10, 30*), linkage analysis (*3–5, 31*), and long-range restriction mapping with pulsed field gel electrophoresis (*8, 11, 12*). Through the cooperation of patient families, clinicians, and CF researchers throughout the world, many families, especially those in whom recombination

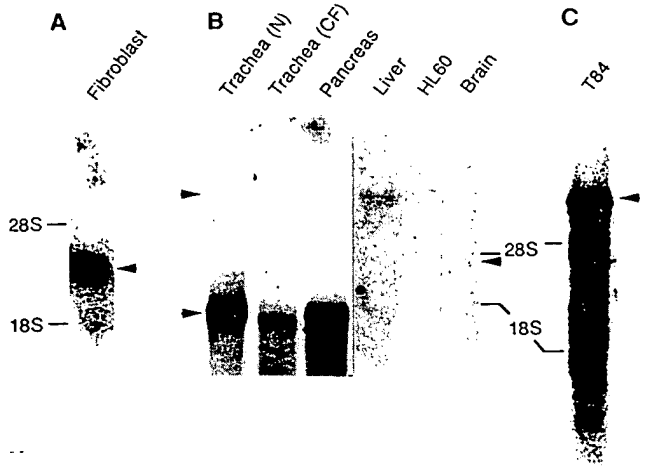


Fig. 4. RNA gel-blot hybridization analysis. RNA hybridization results are shown for the genomic probes G-2 (**A**) and CF16 (**B**). The cDNA clone 10-1 is the probe in (**C**). Approximately 10 μg of total RNA from each human tissue as indicated was separated on a 1 percent formaldehyde gel (*18*). Positions of the 28*S* and 18*S* ribosomal RNA bands are indicated, and arrows indicate the positions of transcripts. HL60 is a human promyelocytic leukemia cell line (*38*), and T84 is a human colon cancer cell line (*39*). Normal (N) and CF trachea are shown.

events had occurred near *CF* (*15, 24*), were identified and made available for genetic mapping studies, which led to an accurate localization of *CF* with respect to the flanking DNA markers (*4, 5, 10, 16*).

In the absence of useful cytogenetic landmarks to pinpoint *CF*, a systematic search of gene sequences within the entire region suggested by genetic data was required. Cloning from pulsed field gels and isolation of undermethylated CpG-rich regions (*7, 8*) served to identify further regions of interest, but cloning of a large contiguous stretch of DNA, as described in this article, allowed a more thorough examination of the region for candidate gene sequences. In this regard, the combination of chromosome walking and jumping appeared to be a highly productive strategy in covering the *CF* region. The jumping technique was particularly useful in bypassing "unclonable" regions, which are estimated to constitute 5 percent of the human genome (*17*). An alternative to this strategy would be the use of yeast artificial chromosome (YAC) vectors which allow cloning of large DNA fragments in the size ranges of 100 to 1000 kb (*32*); however, the construction of sublibraries with phage or cosmid vectors will probably still be required in order to generate a complete restriction map and identify candidate gene sequences from the YAC clones.

The challenge of identifying all gene sequences in a large DNA segment was also formidable, as no single method was guaranteed to succeed. In view of the experience described above, it would be advisable to attempt a combination of all available methods. Searching for sequence conservation by cross-hybridization is rapid, and the ability to use entire phage or cosmid clones represents a substantial simplification of the screening procedure. However, not all large segments of human DNA could be used in this way; simple repetitive sequences (for example, CA repeats) that are highly abundant in other animal species (*33*) can interfere with hybridization analysis. The evolutionary conservation of the E4.3 and H1.6 fragments, for example, was only apparent when these segments were isolated away from neighboring repeats.

The ultimate task in this type of "reverse genetics" approach (*34*) is to prove the identity of a candidate gene as, by definition, the basic biochemical defect of the disease is unknown. Appropriate tissue distribution and predicted properties of the gene product provide strong supporting evidence; these criteria have now been met for the CF gene, as detailed in the accompanying paper by Riordan *et al.* (*18*). The identification of a specific mutation which is found in affected individuals but never appears in normal chromosomes is much more compelling, and this evidence is presented for CF by Kerem *et al.* (*16*), who have now defined the most common CF mutation. Identification of other CF mutations will provide additional support. Expression of the normal cDNA in CF cells, which should correct the phenotypic chloride channel defect, will represent an important confirmation of the identity of the gene (*35*) and will be useful in the elucidation of the precise molecular pathology of the CF defect.

The large size of the CF gene came as somewhat of a surprise; the absence of apparent genomic rearrangements in CF chromosomes and the evidence indicating a limited number of CF mutations (*4*) might have led to an expectation of a small mutational target. The discovery that the most common CF abnormality gives rise to the loss of a single amino acid residue in a functional domain suggests, however, that the phenotype of CF is likely not due to complete loss of function of the gene product. In this regard, CF may be similar to the sickling disorders, where a very specific subset of mutations in the β-globin gene (β^S and β^C) give rise to an altered protein with unusual behavior (*36*). Complete absence of function of the β-globin gene gives rise to a different phenotype (β°-thalassemia); similarly, homozygous loss of function of the CF protein product may lead to a different phenotype.

In summary, the application of genetic and molecular cloning strategies has allowed the cloning of the cystic fibrosis locus on the basis of its chromosomal location, even without the benefit of genomic rearrangements to point the way. Further improvements in "reverse genetics" technology should facilitate the identification of many more genetic loci of biological and medical importance.

REFERENCES AND NOTES

1. T. F. Boat, M. J. Welsh, A. L. Beaudet, in *The Metabolic Basis of Inherited Disease*, C. L. Scriver, A. L. Beaudet, W. S. Sly, D. Valle, Eds. (McGraw-Hill, New York, ed. 6, 1989), pp. 2649–2680.
2. M. Knowles, J. Gatzy, R. Boucher, *J. Clin. Invest.* **71**, 1410 (1983); P. M. Quinton and J. Bijman, *N. Engl. J. Med.* **308**, 1185 (1983); R. A. Frizzell, G. Rechkemmer, R. L. Shoemaker, *Science* **233**, 558 (1986); M. J. Welsh and C. M. Liedtke, *Nature* **322**, 467 (1986); R. A. Schoumacher *et al.*, *ibid.* **330**, 752 (1987); M. Li *et al.*, *Science* **244**, 1353 (1989); T.-C. Hwang *et al.*, *ibid.*, p. 1351; D. W. Landry *et al.*, *ibid.*, p. 1469.
3. L.-C. Tsui *et al.*, *Science* **230**, 1054 (1985); R. G. Knowlton *et al.*, *Nature* **318**, 380 (1985); R. White *et al.*, *ibid.*, p. 382; B. J. Wainwright *et al.*, *ibid.*, p. 384.
4. For reviews, see M. A. Spence and L.-C. Tsui, *Cytogenet. Cell Genet.* **46**, 170 (1988); M. Dean, *Genomics* **3**, 93 (1988); L.-C. Tsui, *Am. J. Hum. Genet.* **44**, 303 (1989).
5. A. Beaudet *et al.*, *Am. J. Hum. Genet.* **39**, 681 (1986); G. M. Lathrop *et al.*, *ibid.* **42**, 38 (1988).
6. F. S. Collins *et al.*, *Science* **235**, 1046 (1987); M. C. Iannuzzi *et al.*, *Am. J. Hum. Genet.* **44**, 695 (1989).
7. F. Michiels, M. Burmeister, H. Lehrach, *Science* **236**, 1305 (1987).
8. X. Estivill *et al.*, *Nature* **326**, 840 (1987).
9. R. Kaiser *et al.*, *Mol. Biol. Rep.* **12**, 3 (1987).
10. J. M. Rommens *et al.*, *Am. J. Hum. Genet.* **43**, 4 (1988).
11. J. M. Rommens *et al.*, *ibid.*, in press.
12. A.-M. Poustka, H. Lehrach, R. Williamson, G. Bates, *Genomics* **2**, 337 (1988); M. L. Drumm *et al.*, *ibid.*, p. 346.
13. M. Dean *et al.*, in preparation.
14. Twelve genomic DNA libraries were used in this study. This included eight phage libraries, one of which was provided by T. Maniatis [E. F. Fritsch, R. M. Lawn, T. Maniatis, *Cell* **19**, 959 (1980)]; the rest were constructed as part of this work according to procedures described in (*20*). Four phage libraries were cloned in λDASH (Stratagene) and three in λFIX (Stratagene), with vector arms provided by the manufacturer. One λDASH library was constructed from Sau 3A–partially digested DNA from a human-hamster hybrid containing human chromosome 7 (4AF/102/K015) (*10*), and the other libraries from human peripheral blood or lymphoblastoid DNA after partial digestion with Sau 3A or total digestion with Bam HI or Eco RI. To avoid loss of unstable sequences, five of the phage libraries were propagated in the recombination-deficient hosts DB1316 (*recD*$^-$) (*17*), CES200 (*recBC*$^-$) (*17*), or TAP90 [T. A. Patterson and M. Dean, *Nucleic Acids Res.* **15**, 6298 (1987)]. Three cosmid libraries were constructed. In one, the vector pCV108 [Y. F. Lau and Y. W. Kan, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 5225 (1983)] was used to clone partially digested (Sau 3A) DNA from 4AF/102/K015 (*10*). A

second cosmid library was prepared by cloning partially digested (Mbo I) human lymphoblastoid DNA into the vector pWE-IL2R, prepared by inserting the Rous sarcoma virus (RSV) promoter-driven DNA for the interleukin-2 receptor α chain (supplied by M. Fordis and B. Howard) in place of the *neo*-resistance gene of pWE15 [G. M. Wahl *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 2160 (1987)]. An additional partial Mbo I cosmid library was prepared in the vector pWE-IL2R-Sal, created by inserting a Sal I linker into the Bam HI cloning site of pWE-IL2R (M. L. Drumm, unpublished data); this allows the use of the partial fill-in technique to ligate Sal I and Mbo I ends, preventing tandem insertions [E. R. Zabarovsky and R. L. Allikmets, *Gene* **42**, 119 (1986)]. Cosmid libraries were propagated in *Escherichia coli* host strains DH1 or 490A [M. Steinmetz, A. Winoto, K. Minard, L. Hood, *Cell* **28**, 489 (1982)]. Single copy DNA segments (free of repetitive elements) near the ends of each phage or cosmid insert were purified and used as probes for library screening to isolate overlapping DNA fragments by standard procedures (*20, 21*). For each walk step, the identity of the cloned DNA fragment was determined by hybridization with a somatic cell hybrid panel to confirm its chromosomal location, and by restriction mapping and DNA gel-blot analysis to confirm its colinearity with the genome. The chromosome jumping library has been described (*6*). The original library was prepared from a preparative pulsed field gel and was intended to contain partial Eco RI fragments of 70 to 130 kb; subsequent experience with this library (including that reported here) indicates that smaller fragments are also represented, and jumps of 25 to 110 kb have been found. The library was plated on *sup*$^-$ host MC1061 and screened by standard techniques (*20*). Positive clones were subcloned into pBRΔAva, and the beginning and end of the jump were identified by Eco RI and Ava I digestion [F. S. Collins, in *Genome Analysis: A Practical Approach*, K. E. Davies, Ed. (IRL, London, 1988), pp. 73–94]. For each clone, a fragment from the end of the jump was checked to confirm its location on chromosome 7. About 10 percent of the clones in this library arise from noncircular ligations and thus give rise to anomalous jumps (F. S. Collins, unpublished).

15. M. Farrall *et al.*, *Am. J. Hum. Genet.* **43**, 471 (1988).
16. B. Kerem *et al.*, *Science* **245**, 1073 (1989); B. Kerem *et al.*, *Am. J. Hum. Genet.* **44**, 827 (1989).
17. A. R. Wyman, L. B. Wolfe, D. Botstein, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 2880 (1985); K. F. Wertman, A. R. Wyman, D. Botstein, *Gene* **49**, 253 (1986); A. R. Wyman, K. F. Wertman, D. Barker, C. Helms, W. H. Petri, *ibid.*, p. 263.
18. J. R. Riordan *et al.*, *Science* **245**, 1066 (1989).
19. S. M. Arfin, R. E. Circullo, V. F. X. Arredondo, M. Smith, *Somatic Cell Genet.* **9**, 517 (1983).
20. T. Maniatis, E. F. Fritsch, J. Sambrook, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982).
21. DNA probes that gave high background signals could often be used more successfully by annealing the boiled probe with sheared, denatured, placental DNA (250 μg/ml) for 60 minutes before the probe was added to the hybridization bag.
22. For example, localization of the Duchenne muscular dystrophy locus was facilitated by cross-species hybridization, as described in A. P. Monaco *et al.* [*Nature* **323**, 646 (1986)].
23. A. P. Bird, *Nature* **321**, 209 (1986); M. Gardiner-Garden and M. Frommer, *J. Mol. Biol.* **196**, 261 (1987).
24. L.-C. Tsui *et al.*, *Am. J. Hum. Genet.* **39**, 720 (1986); R. White *et al.*, *ibid.*, p. 694; W. Berger *et al.*, *Hum. Genet.* **77**, 197 (1987).
25. The SV40-transformed human fibroblast cDNA library, with 1.4×10^6 independent clones, was provided by P. Berg [H. Okayama and P. Berg, *Mol. Cell. Biol.* **3**, 280 (1983)].
26. B. J. Wainwright *et al.*, *EMBO J.* **7**, 1743 (1988).
27. T. G. Fanning and M. F. Singer, *Biochim. Biophys. Acta* **910**, 203 (1987).
28. J. M. Rommens *et al.*, unpublished data.
29. B. Royer-Pokora *et al.*, *Nature* **322**, 32 (1986); M. Koenig *et al.*, *Cell* **50**, 509 (1987); S. H. Friend *et al.*, *Nature* **323**, 643 (1986); D. C. Page *et al.*, *Cell* **51**, 1091 (1987).
30. I. Bartels *et al.*, *Am. J. Hum. Genet.* **38**, 280 (1986); S. Zengerling *et al.*, *ibid.* **40**, 228 (1987).
31. D. Barker *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8006 (1987).
32. D. T. Burke, G. F. Carle, M. V. Olson, *Science* **236**, 806 (1987); B. H. Brownstein *et al.*, *ibid.* **244**, 1348 (1989).
33. R. Meisfeld, M. Krystal, N. Arnheim, *Nucleic Acids Res.* **9**, 5931 (1981); H. Hamada, M. G. Petrino, T. Kakunaga, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 6465 (1982).
34. S. H. Orkin, *Cell* **47**, 845 (1986); F. Ruddle, *Am. J. Hum. Genet.* **36**, 944 (1984).
35. M. L. Drumm *et al.*, *Am. J. Hum. Genet.* **43**, A182 (1988); A. M. Jetten, J. R. Yankaskas, M. J. Stutts, N. J. Willumsen, R. C. Boucher, *Science* **244**, 1472 (1989).
36. H. F. Bunn and B. G. Forget, *Hemoglobin: Molecular, Genetic and Clinical Aspects* (Saunders, Philadelphia, 1986).
37. X. Estivill *et al.*, *Am. J. Hum. Genet.* **44**, 704 (1989).
38. S. J. Collins, F. W. Ruscetti, R. E. Gallagher, R. C. Gallo, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 2458 (1978).
39. K. Dharmsathaphorn *et al.*, *Am. J. Physiol.* **246**, G204 (1984).

40. We thank N. Alon, M. Collins, B. Gerrard, J. Grim, Z. Grzelczak, S. Lok, D. Markiewicz, K. Murphy, L. Naismith, S. Norton, A. Perry, J. Pinnard, N. Plavsic, A. Saulino, C. Stewart, C. TomHon, Z. Weaver, and J. Zielenski for technical assistance; B. Sandri for preparing the manuscript; and J. Yankaskas and R. Boucher for RNA samples. Supported by NIH grants DK39690 (F.S.C.) and DK34944 (L.-C.T.); the Cystic Fibrosis Foundation and the Canadian Cystic Fibrosis Foundation; a postdoctoral fellowship of the Medical Research Council (MRC) of Canada (J.M.R.); an MRC Scientist award (L.-C.T.); the Deutsche Forschungsgemeinschaft (G.M.); the Sellers Foundation (M.B., J.R.R., and L.-C.T.); the North Dakota Cystic Fibrosis Association (L.-C.T. and M.B.); and federal funds from the Department of Health and Human Services under contract NO1-CO-74102 with Program Resources, Inc. (M.D.). F.S.C. is an Associate Investigator in the Howard Hughes Medical Institute.

7 August 1989; accepted 17 August 1989

Probing the basic defect in cystic fibrosis

Lap-Chee Tsui

The Hospital for Sick Children, University of Toronto, Ontario, Canada

The concurrent developments in electrophysiology studies and the identification of the cystic fibrosis transmembrane conductance regulator (CFTR) gene has provided a unique opportunity to probe the basic cellular defect underlying cystic fibrosis. Various properties of the CFTR protein have been deduced from its primary sequence, the variety of mutations in patients and genotype–phenotype correlations, as well as the results of more recent DNA transfection studies. The most exciting observation is the fact that CFTR acts like a cAMP-regulated Cl^- channel.

Current Opinion in Genetics and Development 1991, 1:4–10

Introduction

Cystic fibrosis (CF) is a common autosomal recessive disorder prevalent in the Caucasian population (for a review see [1]). The gene responsible for this disease has recently been identified through molecular cloning studies based on the precise chromosome localization of the disease locus [2–4]. Currently, the most exciting challenge in CF research, as for any disease in which a genetic defect is found before a clear biochemical description of that defect is available, is to identify of the function or functions of the gene product. Of particular importance to CF, however, will be to understand how the same defect can cause problems in a variety of organs.

The abnormality in mucus secretion resulting in chronic obstructive lung disease and pancreatic enzyme insufficiency, together with the elevated levels of sweat electrolytes, are indicative of a general exocrine malfunction in CF [1]. A defect in the regulation of chloride transport has been the most consistent finding in studies of CF epithelial cells (Fig. 1; reviewed in [5•,6•]). This defect appears to be characterized by the lack of a cAMP-mediated channel response, but there have been different descriptions regarding the properties of the chloride channel involved [7–13]. Early single channel patch-clamp studies suggested that the channel is outwardly rectifying with a conductivity of 26–50 picoSiemens [7,8], whereas the more recent measurements indicate that it may be a small linear channel of 4–8 picoSiemens [11–13].

The concurrent developments in CF electrophysiology studies and the identification of the gene has provided a unique opportunity to probe the basic cellular defect underlying this disease. In the past 18 months there has been significant progress in understanding the relationship between the observed defect in chloride transport and the putative protein product of this gene, named the cystic fibrosis transmembrane conductance regulator (CFTR).

Predicted properties of the encoded protein

From its cDNA sequence, CFTR is predicted to contain 1480 amino acid residues, and to have a relative molecular mass of about 170 000 and a pI of 8–9 [3]. In addition, the protein appears to be internally duplicated: there are two repeated motifs, each consisting of six putative membrane-spanning regions followed by an ATP-binding site denoted as the nucleotide-binding fold (NBF) (Fig. 2). The hydrophilic segment preceding the first transmembrane sequence is probably in the cytoplasm, and the bulk of the protein is presumably also internally located. The largest predicted extracellular loop is between transmembrane segments 7 and 8, and contains two possible N-linked glycosylation sites. These features are reminiscent of those observed for many other membrane-associated transport proteins in prokaryotes and eukaryotes [3,14•].

The R-domain is a highly charged cytoplasmic region linking the protein halves (Fig. 3). This unique segment is not found in related transport proteins, and the domain is operationally defined by a single large exon that encodes this sequence. Of the 241 amino acids encoded in this exon, 69 are polar residues, arranged in alternating clusters of positive and negative charges. In addition, nine of the ten potential sites for phosphorylation by protein kinase A and seven of the putative protein kinase C phosphorylation targets found in CFTR are located in this domain. Thus, the R-domain is thought to play an important role in CFTR function [3].

The presence of ATP-binding site consensus sequences (the Walker A and B motifs) in CFTR suggests that ATP

Abbreviations

CF—cystic fibrosis; CFTR—cystic fibrosis transmembrane conductance regulator; NBF—nucleotide-binding fold; PI—pancreatic insufficient; PS—pancreatic sufficient.

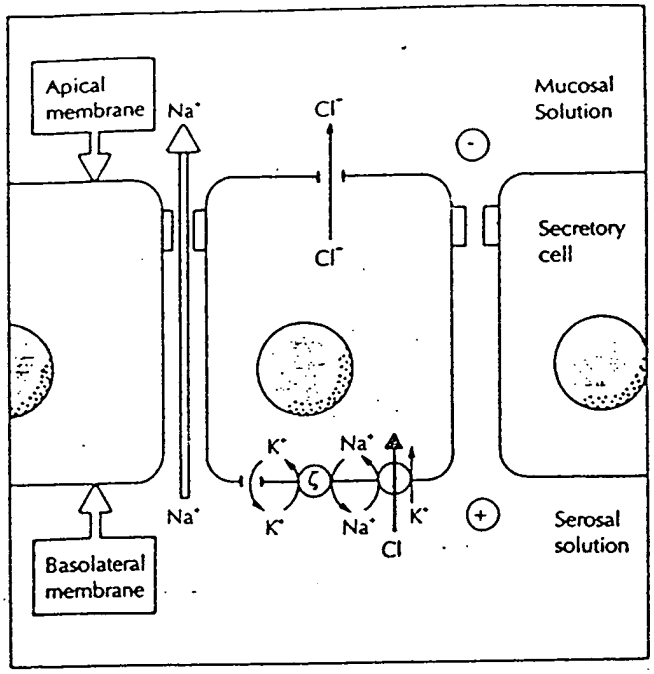


Fig. 1. Cellular model for trans-epithelial chloride secretion (adapted from [38]). Cl^- enters secretory cells across the basolateral membrane, coupled to the entry of Na^+ and K^+ via a loop diuretic-sensitive Na^+–K^+–$2Cl^-$ cotransport process. Sodium is returned to the serosal (plasma-facing) solution by the ouabain-sensitive Na^+–K^+ pump, which also brings K^+ into the cell. Cellular K^+ is returned to the serosal solution via basolateral K^+ channels. These three processes lead to cellular Cl^- accumulation at levels exceeding the electrochemical potential of extracellular Cl^-. Cl^- then exits across the apical membrane by diffusion through Cl^--sensitive channels. Na^+ accompanies Cl^- to the luminal solution by diffusing across the paracellular pathway, driven by the trans-epithelial voltage arising from electrogenic Cl^- transport. The defect in CF cells is thought to be at the point of regulation of Cl^--sensitive channels.

binding (and possibly its hydrolysis) is necessary for the function of the protein [3,14•]. The first experimental evidence for the involvement of ATP in CFTR function came from a study by Thomas *et al.* [15•], who demonstrated binding of ATP to a 67-residue synthetic peptide spanning the central region of the first NBF. This peptide contains the Walker A motif and the central region of the consensus ATP-binding site, but not the B motif; nevertheless, this does not exclude involvement of the B motif in the native protein.

Based on the primary sequence and its structural resemblance to other transport proteins, it is intuitive to suggest that CFTR is an epithelial cell-specific transporter. For example, Hyde *et al.* [14•] argue the following: channels, unlike transporters, do not require ATP hydrolysis, whereas CFTR has two NBFs; ion flow through channels is bidirectional, whereas CFTR resembles a family of unidirectional transporters; a Cl^- channel is expected to have a turnover rate far more rapid than that of transporters; Cl^- channels appear to have varying properties in different tissues, whereas CF affects a variety of tissues; and finally, the molecular weight of a putative Cl^- channel is much less than that of CFTR. Ringe and Pestko [16] further speculate that the products of the arachidonic acid pathway, leukotrienes and prostaglandins, are possible substrates for CFTR. Although these are attractive hypotheses, a different conclusion has been derived from cDNA transfection studies (see below).

Identification of naturally occurring mutations

Useful information about the structure and function of CFTR may be derived from the characterization of both naturally occurring mutations and normal sequence variants. An international consortium (the CF Genetic Analysis Consortium), which consists of 86 research groups from more than 20 countries, has been established to collect this information. To date, over 75 different disease-causing mutations have been reported.

The most common CF mutation (ΔPhe508, or ΔF508 in one-letter code), which accounts for 68% of the global defective gene pool [4,17], is a 3 bp deletion located in exon 10. The primary consequence of this mutation is the deletion of a single phenylalanine residue at position 508, which lies within the first NBF of the predicted polypeptide between the consensus Walker A and B motifs. It has been suggested that ΔPhe508 may prevent proper binding or hydrolysis of ATP, or cause a conformational change affecting CFTR activity [3].

The structural and functional importance of the region surrounding Phe508 has been further elucidated by several additional observations. First, a second 3 bp deletion has been found in the vicinity of ΔPhe508 [18]. This mutation, which is the only other single amino acid deletion found in CFTR so far, deletes the isoleucine residue at position 506 or 507 (hence denoted as Δ1le507). Second, amino acid substitutions for Phe508 and Ile506 have been found in two rare variant alleles that are apparently normal (with either a cysteine residue at position 508 or a valine residue at 506) [19]. Third, direct analysis of the physical properties of a synthetic peptide containing these amino acid residues suggests that deletion of any of the residues could induce a significant structural change in the predicted β-sheet structure and also affect ATP binding [15•]. These observations are consistent, therefore, with the hypothesis that the length of the polypeptide is more important than the identity of the actual amino acid residues in this region [18].

A number of missense mutations have been found within the two NBFs, some of them affecting the highly conserved residues in the Walker motifs (e.g. Gly458 [20], Ser549 [18], and Gly551 [18,21]. These mutations thus confirm that ATP binding (and possibly hydrolysis) is essential for CFTR function. Those mutations that affect non-conserved amino acids (e.g. Asp110→His, Arg117→His, Arg347→Pro, Ala559→Thr, Tyr563→Asn, and Pro574→His [18,21,22]) are perhaps more revealing because the importance of these residues could not otherwise be easily recognized. Understanding the role of these missense mutations may be more difficult at the present time, however, as the three-dimensional structure of the protein is unknown.

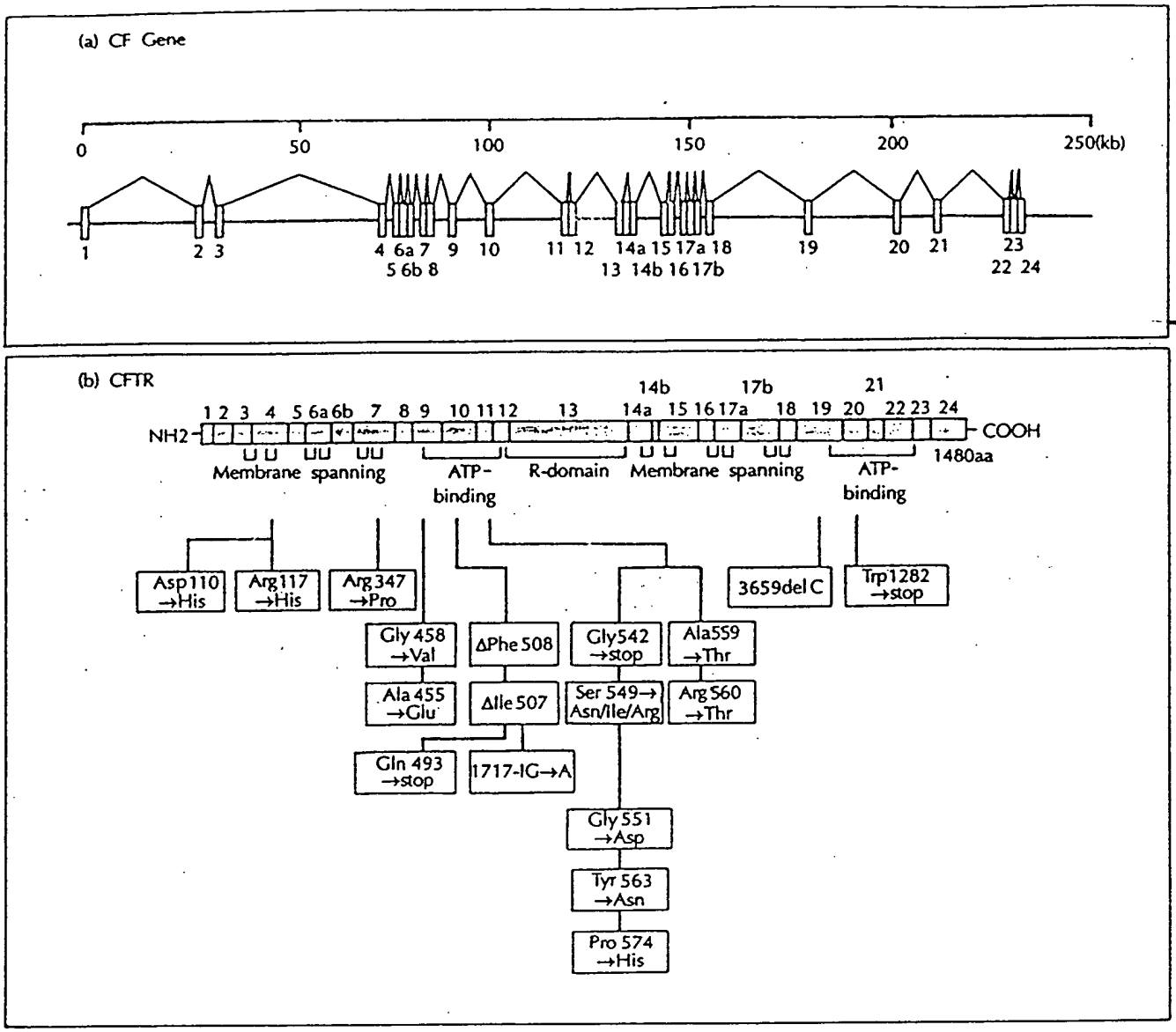


Fig. 2. Schematic diagram (adapted from [39]) of the exon–intron organization of the CFTR gene (a), and the putative domain-type structure of its product (b). The gene is composed of 27 exons. The nomenclature 6a, 6b, 14a, 14b, 17a and 17b reflects intron sequences not originally reported [2]. The positions of the CF mutations, including the major mutation ΔPhe508, are indicated.

The mutation screening data also confirm that the predicted initiation codon [3] is correct; thus, in addition to Asp110→His and Arg117→His [22], there are more than ten other mutations known to precede the next available methionine codon, located in exon 4 (amino acid position 150), indicating that the sequence between exons 1 and 4 is essential for CFTR function.

Approximately half of the known CF mutations are nonsense, frameshift or mRNA splicing mutations (unpublished data reported to the CF Genetic Analysis consortium) [18,20–25]. Although mechanisms such as alternative splicing possibly prevent premature termination of CFTR synthesis for some of these mutant alleles, it is more probable that most, if not all, will produce a null phenotype. Although most of these mutations are detected in compound heterozygotes because of their relatively low individual frequency, there are some examples of CF patients homozygous for such CF mutations who are thought to have a null CFTR phenotype [20,23]. These observations suggest that total absence of CFTR is not lethal and, moreover, patients with a presumed null CFTR phenotype can have merely a mild clinical course of CF [20,23].

Despite extensive searches, only two frameshift mutations have been found towards the end of the R-domain [24,25]. In addition, no sequence polymorphism has been noted in this region among the large number of normal and CF alleles examined (the CF Genetic Analysis Consortium) [18]. The significance of these observations is unknown, but they are consistent with the assumption that the R-domain is essential for CFTR function.

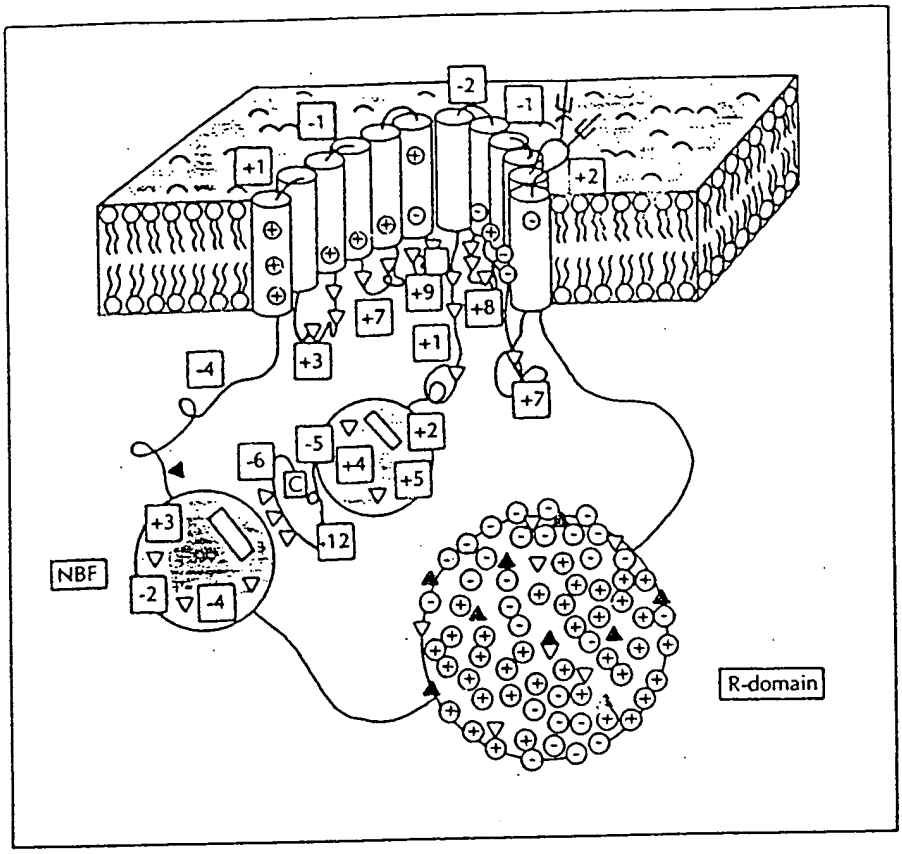


Fig. 3. Schematic model of CFTR (adapted from [3]). The six predicted membrane-spanning helices in each half of the molecule are depicted as cylinders. The position at which ATP enters each of the cytoplasmically-oriented nucleotide-binding folds (NBFs) is marked. The large polar R-domain, which links the two halves, is represented by a sphere. Individual charged amino acid residues are shown as small circles containing the appropriate charge sign. Net charges on the internal and external loops joining the membrane cylinders and on regions of the NBFs are written in boxes. Potential sites for phosphorylation by protein kinase A (▲) or C (▽) and N-glycosylation (ψ) are indicated. The symbol ⊕ denotes the amino acid residue Lys, Arg or His, and the symbol ⊖ denotes the amino acid residue Asp or Glu.

Genotype–phenotype correlation

The remarkable concordance of pancreatic involvement among CF patients of the same family has provided the first suggestion that mutations in the CF gene have a direct clinical consequence [26]. Kerem and coworkers [4] have proposed that there are 'severe' and 'mild' mutations that determine the level of pancreatic function. Approximately 90% of CF chromosomes are believed to carry a severe allele, such as ΔPhe508, whereas the rest carry a mild allele. About 85% of patients, who are pancreatic insufficient (PI), are thought to have two severe alleles; the other 15% of CF patients, who are pancreatic sufficient (PS), are thought to have at least one mild allele. Not all homozygous ΔPhe508 patients are diagnosed as being PI, however. These exceptions may constitute the 10–20% of PS patients whose pancreatic function deteriorates at later stage [27].

Besides ΔPhe508, at least 10 other mutations (Glu493→stop, ΔIle507, 1717-1G→1A, Gly542→stop, Ser549→Ile, Ser549→Arg, Gly551→Asp, Arg560→Thr, 3659delC and Trp1282→stop) are classified as severe alleles [18,20,28]. Only three mild alleles (Ala455→Glu, Tyr563→Asn and Pro574→His) have been defined so far [18]. It is of interest to note that the severe alleles are generally not expected to produce complete CFTR, whereas the mild alleles are missense mutations.

Although most PS patients tend to have better lung function than PI patients, there does not seem to be a direct correlation between genotype and pulmonary phenotype [29•]. Thus, additional genetic and environmental factors are clearly involved in determining lung function, and it is also possible that the role of CFTR is different in different organs. A better understanding of the phenotypes awaits a more detailed analysis of the function of CFTR itself.

Construction of a full-length CFTR cDNA

The cloning of the CFTR gene offers the potential to produce the protein in various expression systems for further biochemical and physiological analyses. Because of the large size of the gene (250 kb) and the lack of knowledge about its promoter region, most studies have used a full-length cDNA insert expressed under the control of a heterologous promoter. The reconstruction of a full-length cDNA from the previously isolated overlapping fragments [3] has been hampered, however, by the instability of a DNA sequence within the coding region (JM Rommens, S Dho, CE Bea, N Kartner, D Kennedy, JR Riordan, L-C Tsui and JK Foshett, unpublished data) [31••,32••]. This technical difficulty was eventually overcome by the use of a low-copy number bacterial plasmid [30••] and, independently, by introducing nucleotide substitutions within the presumptive unstable region [31••–33••]. Such constructs were used for most of the transfection experiments discussed below.

Detection of CFTR protein

A set of specific antibodies directed against CFTR would be extremely valuable in the study of this protein. Unfortunately, cross-reactive materials of unknown nature are also detected by many of the antibodies directed against synthetic peptides or fusion proteins, even in cells that do not appear to express CFTR (on the basis of RNA detection). Nevertheless, with proper controls, and through the use of heterologous transfection systems, Gregory *et al.* [30••] and Kartner *et al.* [32••] have shown that mature CFTR corresponds to a diffuse band of 150–170 kD on SDS–polyacrylamide gels. The proteins synthesized using full-length CFTR transcripts *in vitro* and in expression systems lacking glycosylation activity have a much faster mobility (130–140 kD) [30••,32••]. The difference in mobility is, therefore, probably a result of glycosylation in the mature protein, although the apparent molecular weight of the unglycosylated form is much lower than that predicted from the amino acid sequence. The cause of this aberrant mobility is unknown, but the observation appears to be common among membrane-associated proteins.

The problem of glycosylation has also been studied by Cheng *et al.* [33••] by expressing a number of mutant CFTR variations in monkey COS-7 cells. These transfection studies show that glycosylation is incomplete for a number of mutant proteins (e.g. ΔPhe508, ΔIle507, Lys464→Met, Phe508→Arg and Ser549→Ile). It is suggested that these mutant proteins are recognized as abnormal and thus retained in the endoplasmic reticulm, where they are subsequently degraded. These investigators further propose that the molecular basis of most CF is the absence of mature CFTR at the correct cellular location [33••]. This model, although appealingly straightforward, cannot account for those mutants in which glycosylation appears to be normal (e.g. Arg334→Trp, Gly551→Asp and Lys1250→Met). It may be the case that mutant proteins do predispose slightly different clinical phenotypes depending on whether glycosylation is normal or not, but all the mutants mentioned above appear to be in the severe class, irrespective of glycosylation status. Moreover, it is unclear if glycosylation is required for function [32••]. Therefore, the value of the COS cell transfection system needs further consideration.

Detection of cAMP-regulated chloride conductance

Two independent studies have recently demonstrated that introduction of a CFTR cDNA clone into cells derived from CF patients can confer cAMP-inducible Cl^- permeability like that present in normal cells and absent in CF cells [31••,34••].

In one study, Rich *et al.* [34••] introduced a CFTR-expressing clone into primary and transformed CF airway epithelial cells, using the vaccinia virus expression system. Cl^- channel activity, as measured by iodide flux, was monitored in a single cell assay system in which the transfected cells were loaded with the halide-sensitive dye SPQ (6-methyl-N-(3-sulfopropyl)quinolinium). Upon induction of intracellular cAMP levels with forskolin, increase of I^- flux was more rapid in cells transfected with the normal cDNA than in those transfected with the variant containing the ΔPhe508 mutation. There was considerable cell to cell variability in the response, however. Nevertheless, patch clamp measurements show that cAMP-induced whole cell current were significantly higher in cells with the normal CFTR than in those with the ΔPhe508 variant.

In the other study, Drumm *et al.* [31••] used the retrovirus-mediated gene transfer method to introduce a CFTR-expressing clone into a CF pancreatic adenocarcinoma cell line, CF-PAC. Although clonal variation was also observed, a significant increase of ^{125}I efflux in the cells containing intact CFTR was observed upon forskolin treatment. In addition, a substantial increase in whole cell current was observed only in cells expressing intact CFTR, whereas no increase in current was detectable in cells containing the control vector. The current–voltage relationship measured for the activated channel is indicative of a linear channel with selectivity for anions.

Although studies described above serve to confirm the identity of the gene encoding CFTR and to establish the necessary first step in developing gene therapy strategies for treating the disease, they provide no further information about the function of the protein. It is unclear from these experiments whether the appearance of cAMP-regulated Cl^- conductance in the CF epithelial cells is a result of the restoration of regulation of existing channels, or the introduction of new regulatable channels. An important insight into this question was provided, however, by a series of transfection studies with non-epithelial cell types (JM Rommens, S Dho, CE Bear, N Kartner, D Kennedy, JR Riordan, L-C Tsui and JK Foskett, unpublished data) [32••,35••].

Using the vaccinia virus system, Anderson *et al.* [35••] transfected the CFTR gene into HeLa, Chinese hamster ovary and NIH 3T3 fibroblast cells, and detected an increased anion permeability and chloride current in the presence of cAMP. Similar increases were not observed for the non-transfected control or for cells expressing a mutant CFTR (ΔPhe508). The cAMP-activated Cl^- currents were comparable to those seen in epithelial cells expressing endogenous CFTR or CF cells expressing exogenous CFTR. Comparable observations were made by Kartner *et al.* [32••] for insect cells (Sf9) transiently transfected with a baculovirus construct containing a full length CFTR cDNA insert. The current–voltage relationship for the cAMP-induced current also appears to be linear, and the single channel conductance is low (8 picoSiemens). We have demonstrated that it is possible to establish stable mouse fibroblast (L) cell lines expressing CFTR and to detect the same cAMP-regulated Cl^- channel activity. Furthermore, the levels of CFTR expression and activity in the transfected L cells are highly reproducible, and are comparable to those observed in human epithelial cell lines.

Results from the transfection studies in non-epithelial cells strongly suggest, therefore, that CFTR itself is a regulated Cl^- channel. Although the formal possibility that CFTR merely serves to regulate latent Cl^- channels in the transfected cells cannot be completely excluded, the presence of channels with the same properties in different cells types from different species argues against this interpretation. On the other hand, it remains to be explained how CFTR acts as a Cl^- channel, especially as the primary structure of CFTR is more suggestive of an active transporter than a channel. Can the R-domain perform the necessary magic? In addition, the current–voltage relationship measured for the CFTR channel is apparently linear, whereas the previous single channel data showed that the defective regulation observed in CF cells affected an outwardly rectifying Cl^- channel. It is also necessary to explain a number of other abnormalities observed in CF cells, such as in Na^+ transport [36] and sulfation [37].

Perspective

It is possible that most of the observed defects in CF cells are a result of secondary consequences. A compromise suggestion is that CFTR is a multifunctional protein, acting as a channel as well as being capable of regulating other cellular activities. Judging by the speed of research in CF nowadays, answers to these questions will probably reach the press before this article leaves the editor's desk.

References and recommended reading

Papers of special interest, published within the annual period of review, have been highlighted as:

- • of interest
- •• of outstanding interest

1. BOAT TF, WELSH MJ, BEAUDET AL: Cystic Fibrosis. In *The Metabolic Basis of Inherited Disease (6th edn)* edited by Scriver CR, Beaudet AL, Sly WS, Valle D [book]. New York: McGraw-Hill, 1989, pp. 2649–2680.

2. ROMMENS JM, IANNUZZI MC, KEREM B, DRUMM ML, MELMER G, DEAN M, ROZMAHEL R, COLE JL, KENNEDY D, HIDAKA N, *ET AL*: Identification of the Cystic Fibrosis Gene: Chromosome Walking and Jumping. *Science* 1989, 245:1059–1065.

3. RIORDAN JR, ROMMENS JM, KEREM B, ALON N, ROZMAHEL R, GRZELCHAK Z, ZIELENSKI J, LOK S, PLAVSIC N, CHOU J-L *ET AL*: Identification of the Cystic Fibrosis Gene: Cloning and Characterization of Complementary DNA. *Science* 1989, 245:1066–1073.

4. KEREM B, ROMMENS JM, BUCHANAN JA, MARKIEWICZ D, COX TK, CHAKRAVARTI A, BUCHWALD M, TSUI L-C: Identification of the Cystic Fibrosis Gene: Genetic Analysis. *Science* 1989, 245:1073–1080.

5. • QUINTON PM: Cystic Fibrosis: a Disease in Electrolyte Transport. *FASEB J* 1990, 4:2709–2717.

A review of electrolyte transport in CF patients.

6. • WELSH MJ: Abnormal Regulation of Ion Channels in Cystic Fibrosis Epithelia. *FASEB J* 1990, 4:2718–2725.

A review of the electrophysiology of CF epithelial cells.

7. WELSH MJ, LIEDKE CM: Chloride and Potassium Channels in Cystic Fibrosis Airway Epithelia. *Nature* 1986, 322:467–470.

8. FRIZZELL RA, RECHKEMMER G, SHOEMAKER RL: Altered Regulation of Airway Epithelial Cell Chloride Channels in Cystic Fibrosis. *Science* 1986, 233:558–560.

9. HWANG T-C, LU L, ZEITLIN PL, GRUENERT DC, HUGANIR AR, GUGGINO WB: Cl^- Channels in CF: Lack of Activation by Protein Kinase C and cAMP-Dependent Protein Kinase. *Science* 1989, 244:1351–1353.

10. LI M, MCCANN JD, ANDERSON MP, CLANCY JP, LIEDKE CM, NAIRN AC, GREENGARD P, WELSH MJ: Regulation of Chloride Channels by Protein Kinase C in Normal and Cystic Fibrosis Airway Epithelia. *Science* 1989, 244:1353–1356.

11. GRAY MA, HARRIS A, COLEMAN L, GREENWELL JR, ARGENT BE: Two Types of Chloride Channels on Duct Cells Cultured from Human Fetal Pancreas. *Am J Physiol* 1989, 257:C240–C251.

12. CLIFF WH, FRIZZELL RA: Separate Cl^- Conductances Activated by cAMP and Ca^{2+} in Cl^- Secreting Epithelial Cells. *Proc Natl Acad Sci USA* 1990, 87:4956–4960.

13. TABCHARANI JA, LOW W, ELIE D, HANRAHAN JW: Low-Conductance Chloride Channel Activated by cAMP in the Epithelial Cell Line T84. *FEBS Letters* 1990, 270:157–164.

14. • HYDE SC, EMSLEY P, HARTSHORN MJ, MIMMACK MM, GILEADI U, PEARCE SR, GALLAGHER MP, GILL DR, HUBBARD RE, HIGGINS CF: Structural Model of ATP-binding Proteins Associated with Cystic Fibrosis, Multidrug Resistance and Bacterial Transport. *Nature* 1990, 346:362–365.

A tertiary structure model of the ATP-binding domain of a class of transport proteins is presented. The modeling is based on similarities between the predicted secondary structures of members of this family and the previously determined structure of adenylate kinase. The authors include CFTR in this class of proteins, and argue that CFTR is probably not a chloride channel.

15. •• THOMAS PJ, SHENBAGMURTHI P, YSERN X, PEDERSEN PL: Cystic Fibrosis Transmembrane Conductance Regulator: Nucleotide Binding to a Synthetic Peptide. *Science* 1991, 251:255–257.

A peptide 67 amino acids in length, corresponding to the central region of the nucleotide-binding fold of CFTR, has been chemically synthesized and purified. This peptide is capable of binding adenine nucleotides, with the binding of ATP having the highest affinity. Circular dichrosim spectroscopy shows that this peptide assumes a predominantly β-sheet structure, conformation of which can be significantly altered by deletion of Phe508.

16. RINGE D, PIETKO GA: Cystic Fibrosis: A Transport Problem? *Nature* 1990, 346:312–313.

17. THE CYSTIC FIBROSIS GENETIC ANALYSIS CONSORTIUM: Worldwide Survey of the ΔF508 Mutation — Report from the Cystic Fibrosis Genetic Analysis Consortium. *Am J Hum Genet* 1990, 47:354–359.

18. KEREM B, ZIELENSKI J, MARKIEWICZ D, BOZON D, GAZIT E, YAHAV J, KENNEDY D, RIORDAN JR, COLLINS FS, ROMMENS JR, TSUI L-C: Identification of Mutations in Regions Corresponding to the Two Putative Nucleotide (ATP)-Binding Folds of the Cystic Fibrosis Gene. *Proc Natl Acad Sci USA* 1990, 87:8447–8451.

19. KOBAYASHI K, KNOWLES M, O'BRIEN WE, BEAUDET AL: Benign Missense Variations in the Cystic Fibrosis Gene. *Am J Hum Genet* 1990, 47:611–615.

20. CUPPENS H, MARYSEN P, DE BOECK C, CASSIMAN JJ: Study of the G542X and G458V Mutations in a Sample of Belgian Patients. *Pediatr Pulmonol* 1990, 5 (Suppl):205.

21. CUTTING GR, KASCH LM, ROSENSTEIN BJ, ZIELENSKI J, TSUI L-C, ANTONARAKIS SE, KAZAZIAN HH JR: A Cluster of Cystic Fibrosis Mutations in the First Nucleotide-Binding Fold of the Cys-

tic Fibrosis Conductance Regulator Protein. *Nature* 1990, 346:366–369.

22. DEAN M, WHITE M, AMOS J, GERRARD B, STEWART C, KHAW K-T, LEPPERT M: Multiple Mutations in Highly Conserved Residues are Found in Mildly Affected Cystic Fibrosis Patients. *Cell* 1990, 61:863–870.

23. CUTTING GR, KASCH LM, ROSENSTEIN BJ, TSUI L-C, KAZAZIAN HH JR, ANTONARAKIS SE: Two Patients with Cystic Fibrosis, Nonsense Mutations in Each Cystic Fibrosis Gene, and Mild Pulmonary Disease. *New Eng J Med* 1990, 323:1685–1689.

24. ZIELENSKI J, BOZON D, KEREM B, MARKIEWICZ D, DURIE P, ROMMENS JM, TSUI L-C: Identification of Mutations in Exons 1 Through 8 of the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) Gene. *Genomics* 1991, 10, in press.

25. WHITE MB, AMOS J, HSU JMC, GERRARD B, FINN P, DEAN M: A Frame-Shift Mutation in the Cystic Fibrosis Gene. *Nature* 1990, 344:665–667.

26. COREY M, DURIE P, MOORE D, FORSTNER GG, LEVISON H: Familial Concordance of Pancreatic Function in Cystic Fibrosis. *J Pediatr* 1989, 115:274–279.

27. WATERS DL, DORNEY SFA, GASKIN KJ, GRUCA MA, O'GHALLORAN M, WILCKEN B: Pancreatic Function in Infants Identified as Having Cystic Fibrosis in a Neonatal Screening Program. *New Engl J Med* 1990, 322:303–308.

28. VIDAUD M, FANEN P, MARTIN J, GHANEM N, NICOLAS S, GOOSSENS M: Three Mutations in the CFTR Gene in French Cystic Fibrosis Patients: Identification by Denaturing Gradient Gel Electrophoresis. *Hum Genet* 1990, 85:446–449.

29. • KEREM E, COREY M, KEREM B, ROMMENS J, MARKIEWICZ D, LEVISON H, TSUI L-C, DURIE P: Association Between the ΔF508 Mutation and Phenotypes in Cystic Fibrosis. *New Engl J Med* 1990, 323:1517–1522.

This paper describes a correlation study between the genotypes and phenotypes of 293 CF patients. The data suggest that the variable clinical course in CF patients can be attributed to specific genotypes at the CF locus.

30. •• GREGORY RJ, CHENG SH, RICH DR, MARSHALL J, PAUL S, HEHIR K, OSTEDGAARD L, KLINGER KW, WELSH MJ, SMITH AE: Expression and Characterization of the Cystic Fibrosis Transmembrane Conductance Regulator. *Nature* 1990, 347:382–386.

CFTR synthesis is monitored in a reticulocyte translation system and in a transfection assay based on the vaccinia virus–T7 polymerase hybrid expression system. These studies show that CFTR normally exists in a glycosylated form in mammalian cells.

31. •• DRUMM ML, POPE HA, CLIFF WH, ROMMENS JM, MARVIN SA, TSUI L-C, COLLINS FS, FRIZZELL RA, WILSON JM: Correction of the Cystic Fibrosis Defect *In Vitro* by Retrovirus-Mediated Gene Transfer. *Cell* 1990, 62:1227–1233.

The retrovirus-mediated gene transfer procedure is used to introduce the full length CFTR cDNA into a pancreatic adenocarcinoma cell line derived from a CF patient. This cell line shows the Cl^- transport abnormalities characteristic of CF but, upon acquisition of the CFTR cDNA, a cAMP-inducible chloride permeability can be detected. This result indicates that expression of the normal CFTR gene confers cAMP-dependent chloride channel regulation on CF epithelial cells.

32. •• KARTNER N, HANRAHAN JW, JENSEN TJ, NAISMITH AL, SUN S, ACKERLEY CA, REYES EF, TSUI L-C, ROMMENS JR, BEAR CE, RIORDAN JR: Expression of the Cystic Fibrosis Gene in Non-Epithelial Invertebrate Cells Produces a Regulated Anion Conductance. *Cell* 1991, 64:681–691.

The baculovirus expression system is used to express CFTR in Sf9 insect cells. Concomitant with CFTR expression, these cells exhibit a new cAMP-stimulated anion permeability. The conductance resembles that present in several CFTR expressing human epithelial cells. This finding demonstrates that CFTR can function in heterologous non-epithelial cells, and lends support to the possibility that CFTR may itself be a regulated anion channel.

33. •• CHENG SH, GREGORY RJ, MARSHALL J, PAUL S, SOUZA DW, WHITE GA, O'RIORDAN CR, SMITH AE: Defective Intracellular Transport and Processing of CFTR is the Molecular Basis of Most Cystic Fibrosis. *Cell* 1990, 63:827–834.

Mutant CFTR constructs are introduced into monkey COS-7 cells. The study shows that mature, fully glycosylated CFTR is absent from cells containing a subset of the mutant constructs. Instead, an incompletely glycosylated version of the protein is detected. The authors propose that the mutant proteins are recognized as abnormal and remain incompletely processed in the endoplasmic reticulum, where they are subsequently degraded. The glycosylation sites previously predicted [3] appear to be used.

34. •• RICH DR, ANDERSON MP, GREGORY RJ, CHENG SH, PAUL S, JEFFERSON DM, MCCANN JDM, KLINGER KW, SMITH AE, WELSH MJ: Expression of Cystic Fibrosis Transmembrane Conductance Regulator Corrects Defective Chloride Channel Regulation in Cystic Fibrosis Airway Epithelial Cells. *Nature* 1990, 347:358–363.

The vaccinia virus expression system is used to introduce the full-length CFTR cDNA into primary and transformed respiratory cells derived from CF patients. Iodide flux and patch-clamp measurements show that cells transfected with a normal CFTR but not the ΔPhe508 cDNA have a more rapid iodide flux in the presence of forskolin. The results are taken to mean that expression of CFTR corrects the defective Cl^- channel regulation in these cells.

35. •• ANDERSON MP, RICH DP, GREGORY RJ, SMITH AE, WELSH MJ: Generation of cAMP-Activated Chloride Currents by Expression of CFTR. *Science* 1991, 251:679–682.

The CFTR gene was introduced into HeLa, Chinese hamster ovary and NIH 3T3 mouse fibroblast cells using the vaccinia virus transfection and expression system. Increases in anion permeability and chloride current were detected in CFTR-expressing cells treated with cAMP, but not in cells expressing a mutant protein or in non-transfected cells. The simplest interpretation of these observations is that CFTR is itself a cAMP-activated chloride channel.

36. BOUCHER RC, STUTTS MJ, KNOWLES MR, CANTLEY L, GATZY JT: Na^+ Transport in Cystic Fibrosis Respiratory Epithelia. Abnormal Basal Rate and Response to Adenylate Cyclase Activation. *J Clin Invest* 1986, 78:1245–1252.

37. CHENG P-W, BOAT TF, CRANFILL K, YANKASKAS JR, BOUCHER RC: Increased Sulfation of Glycoconjugates by Cultured Nasal Epithelial Cells from Patients with Cystic Fibrosis. *J Clin Invest* 1989, 84:68–72.

38. FRIZZELL RA, HALM DR: Chloride Channel in Epithelial Cells. *Curr Top Membr Transport* 1990, 37:247–282.

39. ZIELENSKI J, ROZMAHEL R, BOZON D, KEREM B, GRZELCZAK Z, RIORDAN JR, ROMMENS JM, TAUI L-C: Genomic DNA Sequence of the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) Gene. *Genomics* 1991, 10, in press.

L-C Tsui, Department of Genetics, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, M5G 1X8, Canada.

ON REQUEST FOR RECONSIDERATION

January 17, 1989

Appellants request us to reconsider our holding mailed August 17, 1988 in which we affirmed the examiner's decision rejecting claims 1, 11, 12, 17 and 18.

At page 7 of their request, appellants state that there is no basis in the art of record for reasonably predicting that human beta-NGF could be produced by recombinant host cells. However, appellants appear to misapprehend the basis for our decision. It is our explicit holding that the product to which appellants' claims are directed would have been expected to be the same or substantially the same as that of the human nerve growth factor isolated by Goldstein and by Walker. At page 180 of his article, Goldstein states, under "DISCUSSION" that "we have demonstrated that human placental cotyledons are a suitable source for the purification of human NGF." Likewise, at page 195 of the other publication item cited by the examiner, Walker, in the "Summary", reports "Human B-nerve growth factor (hNGF) was purified from term human placenta." In the last six lines of the first paragraph at page 195, Walker discloses that:

> Recently, Goldtein and coworkers (14) isolated and purified the biologically active β subunit of NGF from term human placenta. The present report confirms the presence of human β–NGF (hNGF) in term human placenta and reports the lack of immonocross-reactivity between mouse (mNGF) and hNGF using 6 different antisera to mNGF.

Beginning at page 5 of our decision, we pointed out that the legal principles enunciated in cases involving product-by-process claims are considered to be applicable herein. In support thereof, we cited the decision in *In re Brown*, which clearly explains the basis for our holding. That is, where the product disclosed in the prior art reasonably appears to be either identical with or slightly different from a product claimed by an applicant, there is pragmatic justification for placing the burden of going forward on the applicant. Furthermore, at page 10 of our decision, after acknowledging the apparent conflict between our opinion and the court's holding in *In re Wakefield*, we asserted that our decision is consonant with the overwhelming weight of current patent jurisprudence. Nevertheless, at page 16 of the request for reconsideration, appellants contend that they can find *no* decisions that support our position. Accordingly, appellants' attention is invited to the decisions in *In re Thorpe*, 777 F.2d 695, 227 USPQ 964 (Fed. Cir. 1985); *In re Marosi*, 710 F.2d 799, 218 USPQ 289 (Fed. Cir. 1983); *In re Fitzgerald*, 619 F.2d 67, 205 USPQ 594 (CCPA 1980); *In re Wertheim*, 541 F.2d 257, 191 USPQ 90 (CCPA 1976); *In re Avery*, 518 F.2d 1228, 186 USPQ 161 (CCPA 1975); *In re Fessman*, 489 F.2d 742, 180 USPQ 324 (CCPA 1974); and *In re Luck*, 476 F.2d 650, 177 USPQ 523 (CCPA 1973), just to name a few. Additionally, for appellants' convenience, we quote the following passage from *In re Best*, 195 USPQ 433-434, cited at page 7 of our decision:

> Where, as here, the claimed and prior art products are identical or substantially identical, or are produced by identical or substantially identical processes, the PTO can require an applicant to prove that the prior art products do not necessarily or inherently possess the characteristics of his claimed product. See In re Ludtke,[2] supra. Whether the rejection is based on 'inherency' under 35 USC 102, on 'prima facie obviousness' under 35 USC 103, jointly or alternatively, the burden of proof is the same, and its fairness is evidenced by the PTO's inability to manufacture products or to obtain and compare prior art products.

[4] Having disposed of the arguments adverting to the product-by-process rationale, we note that, beginning at page 12 of their petition, appellants contend, in effect, that the pure form of a known substance may be patentable. To support their position, appellants refer to the famous "aspirin" case, cited at page 13 of the petition. Nonetheless, the *mere* purity of a compound, in itself, does not render the substance unobvious. Compare the decisions in *Ex parte Hartop*, 139 USPQ 525 (Bd.App. 1962); *Ex parte Steelmand*, 140 USPQ 189 (Bd.App. 1962); *In re Mehta*, 52 CCPA 1615, 347 F.2d 859, 146 USPQ 284 (1965); and *In re Avery*, *supra*. Furthermore, with respect to the decision in *In re Williams*, cited at page 14 of the petition, appellants' attention is invited to the subsequent decision in *In re Adamson*, 47 CCPA 839, 275 F.2d 952, 125 USPQ 233 (1960), which shows that the *Williams* decision resulted from the absence of relevant available evidence and does not represent a controlling rule of law. Also compare *In re*

[2] 58 CCPA 1159, 441 F.2d 660, 169 USPQ 563 (1971).

Anthony, 56 CCPA 1443, 414 F.2d 1383, 162 USPQ 594 (1969).

[5] At page 11 of our decision, we noted that no objective evidence had been provided establishing that a method was unknown to those skilled in the field whereby the claimed material might have been synthesized. In response thereto, at page 6 of the petition, appellants complain that this improperly shifts the burden of proof to them and places them in the untenable position of having to prove a negative of enormous scope. We disagree. Rather, we are of the opinion that, to raise the question of nonenablement, appellants must, at the very least, provide a declaration by a person having ordinary skill in the subject art that no method was known to him prior to the claimed invention whereby the claimed material might have been synthesized. In this connection, attention is invited to the decision in *In re Collins*, 59 CCPA 1170, 462 F.2d 538, 174 USPQ 333 (1972). It will be noted that in said decision, not only was an affidavit required, but the court agreed with the board that the submitted affidavit failed to establish that there was no known or obvious way to make heat exchangers falling within the scope of the appealed claims. However, in the interest of reducing the issues in this case, we will agree, *arguendo*, that the only methods for obtaining human nerve growth factor, other than that of appellants, are those disclosed by Goldstein and by Walker, of record.

At page 11 of the petition, appellants again rely upon Breakefield as casting doubt on the identity of the materials reported in the cited references. At page 12 of said petition, appellants state that the Breakefield reference "serves as expert testimony". We are then requested to provide our own evidence to counter the Breakefield et al. findings. Nevertheless, we will decline appellants' invitation.

The reason for requiring evidence in declaration or affidavit form is to obtain the assurances that any statements or representations made are correct, as provided by 35 U.S.C. 25 and 18 U.S.C. 1001. To permit the Breakefield publication, coauthored by one of the appellants herein, to substitute for expert testimony would circumvent the guarantees built into the statute by Congress. Accordingly, it is clear that we have no duty to offer evidence to counter the statements made by Breakefield. Rather, we are charged with the obligation of balancing all of the cited evidence of obviousness against the submitted evidence of non-obviousness. See the paragraph bridging pages 7 and 8 of our decision. In so weighing the evidence, we determined that the Breakefield publication item was inadequate to counterbalance the factual findings in the two publication items provided by the examiner. Consequently, we are convinced that our decision in the present case fully complies with the requirements of the statute and 37 C.F.R. 1.196(a).

Focusing now on claims 17 and 18, we observe that appellants, beginning at page 17 of their petition, again separately argue the patentability of human methionine beta-NGF. However, appellants have failed to respond to our finding that the protein containing the terminal methionyl group is substantially identical in structure to that purified by Goldstein and Walker. Since the decisions, such as *In re Brown*, of record, and *In re Best*, of record, agree that the disclosure of a *substantially* identical material in a prior art reference is sufficient to establish a prima facie case of obviousness and shift the burden of proof to appellants, we find no reason to arrive at a different conclusion.

At page 21 of their request, appellants acknowledge that the exhibits accompanying said request are newly cited and have not been considered by the examiner. Nonetheless, appellants request us to consider the references to economize our time. Ignoring the submitted publication items, appellants caution, will insure that said items will be presented in a continuing application. However, the same reasoning might be employed to extend the prosecution in any application handled by an examiner or to dispute any decision rendered by this Board. Since there must be an end to prosecution in any particular case, the mere possibility of further prosecution in a continuing application is insufficient reason for us to consider the publications cited by appellants herein. Compare *In re Fessman*, *supra*.

Although, to the extent indicated, we have reconsidered our decision, we decline to make any changes therein.

DENIED.

APPENDIX

1. Human β-NGF comprising the amino acid sequence ser-ser-ser-his-pro-ile-phe-his-arg-gly-glu-phe-ser-val-cys-asp-ser-val-ser-val-trp-vil-gly-asp-lys-thr-thr-ala-thr-asp-ile-lys-gly-lys-glu-val-met-val-leu- gly-glu-val-asn-ile-asn-asn-ser-val-phe-lys-gln-tyr-phe-phe-glu-thr-lys- cys-arg-asp-pro-asn-pro-val-asp-ser-gly-cys-arg-gly-ile-asp-ser-lys-his-trp-asn-ser-tyr-cys-thr-thr-thr-his-thr-phe-val-lys-ala-leu-thr-met-asp-gly-lys-gln-ala-ala-trp-arg-phe-ile-arg-ile-asp-thr-ala-cys-

val-cys-val-leu-ser-arg-lys-ala-val-arg and which is free of other proteins of human origin.

District Court, N.D. California

Bausch & Lomb Inc. v. Barnes-Hind/Hydrocurve Inc.

No. C 83-20283 RPA

Decided March 22, 1989

PATENTS

1. Patentability/Validity — Obviousness — Relevant prior art (§115.0903)

Relevant prior art for determining obviousness of laser-engraved soft contact lens invention consists primarily of patent for laser apparatus for cutting holes in hard contact lenses and patent disclosing use of laser to engrave plastic surface of printing plate, plus development of laser technology between 1968 and 1976.

2. Patentability/Validity — Obviousness — Combining references (§115.0905)

Patent for laser-engraved soft contact lens is not obvious in view of prior patent on laser apparatus for fenestration of hard contact lenses in combination with prior patent disclosing use of laser to engrave plastic surface of printing plate, since latter patent "taught away" from using process on materials suitable for soft contact lenses, and since Court of Appeals for the Federal Circuit held, as law of case, that one skilled in art would not have construed laser fenestration teachings of former patent as applying to soft contact lenses.

3. Patentability/Validity — Obviousness — Secondary considerations generally (§115.0907)

Evidence of secondary considerations is not persuasive as to non-obviousness of patent for laser-engraved soft contact lens, since evidence of copying is inconclusive, since need to use automatic engraving was not long-felt but rather arose in early 1980s as result of growth in sales of soft contact lenses, and since patented marking system did not bring patent holder commercial success.

4. Infringement — Literal infringement (§120.05)

Plaintiff's patent for laser-engraved soft contact lens is infringed by defendants' contact lenses, since issue of infringement involves only whether surface of defendants' lens surrounding laser marks is "smooth," since specification indicates that "smooth" means absence of ridges that would scratch eye or eyelid, and since defendants' lenses do not inflame or irritate wearers' eyes.

Particular patents — General and mechanical — Contact lenses

4,194,814, Fischer, McCandless and Hager, transparent opthalmic lens having engraved surface indicia, valid and infringed.

On remand from the U.S. Court of Appeals for the Federal Circuit; 230 USPQ 416.

Action by Bausch & Lomb Inc. against Barnes-Hind/Hydrocurve Inc. and Barnes-Hind International Inc., for patent infringement. On remand from decision vacating judgment for defendants. Judgment for plaintiff.

Laurence H. Pretty and Craig S. Summers, of Pretty, Schroeder, Brueggemann & Clark, Los Angeles, Calif.; Anne L. Enea, of Ferrari, Alvarez, Olsen & Ottoboni, San Jose, Calif., for plaintiff.

John M. Calimafde, Paul H. Blaustein, and Dennis J. Mondolino, of Hopgood, Calimafde, Kalil, Blaustein & Judlowe, New York, N.Y.; Douglas K. Tribble, of Pillsbury, Madison & Sutro, San Jose, for defendant.

Aguilar, J.

FINDINGS OF FACT AND CONCLUSIONS OF LAW FOLLOWING REMAND

This patent infringement case returns to this Court on remand from the United States Court of Appeals for the Federal Circuit. The circuit vacated the judgment entered after trial as improper for the following reasons:

(1) this Court did not explicitly set forth in its Order the presumption of validity that it awarded the patent under 35 U.S.C. §282;

(2) the Court did not set forth factual findings on the four inquiries mandated by *Graham v. John Deere Co.*, 383 U.S. 1, 17 [148 USPQ 459, 467] (1966); and

(3) the circuit court found this Court engaged in improper claim construction.

Time has not shown any evidence of actual confusion. However, evidence of actual confusion is not necessary for establishing a claim of likelihood of confusion. *Miles Shoes, Inc. v. R.H. Macy & Co., Inc.*, 199 F.2d 602, 603 [95 USPQ 170, 171-72] (2d Cir. 1952), *cert. denied*, 345 U.S. 909 [96 USPQ 457] (1953); *La Touraine Coffee Co. v. Lorraine Coffee Co.*, 157 F.2d 115, 117 [70 USPQ 429, 431-32] (2d Cir. 1946), *cert. denied*, 329 U.S. 771 [71 USPQ 328] (1946). Moreover, in light of the circumstances of this injunction, this is not fatal to Time's claim of trade dress infringement. Globe has published only one issue of *Celebrity* which was on the stands for only a short period of time before Time brought this action. Given the prior findings of likelihood of confusion, it is understandable that Time would not have been able to gather such information as of the start of this action.

[4] For the reasons set forth above, Time has established a trade dress in the *People* cover format, the similarity of the *Celebrity* cover format, and that Globe's intentional imitation of *People*'s cover format in an effort to capitalize on their reputation and advertising. Further, the magazines are very close in competitive proximity, and the buying habits of consumers lead to a conclusion that there is a likelihood of confusion as to the source of publication of *Celebrity* magazine. Each of these factors weigh in Time's favor. It is fair to conclude therefore, that there is a likelihood that the public will be confused and that Time has proven its cause of action for the infringement of its trade dress.

Irreparable Injury

[5] As indicated earlier to obtain a preliminary injunction, Time must establish that *Celebrity*'s use of its cover format will cause irreparable injury. Where a party seeks a preliminary injunction in a trademark infringement case, irreparable harm is demonstrated by a showing of likelihood of confusion as to the source or sponsorship of the magazine. *Home Box Office v. Showtime/The Movie Channel*, 832 F.2d 1311, at 1314 [4 USPQ2d 1789, 1791]. Here Time has shown a likelihood of confusion sufficient to meet the showing of irreparable harm.

Balance of the Hardships

Globe has printed 125,000 copies of the 1989 issue, approximately half of their regular level of production. However, they have only published one issue with the new cover and there are numerous other cover formats from which they could choose a new one. Further, given the finding of irreparable harm to Time and likelihood of confusion among consumers, allowing Globe to continue to use its current cover format could cause serious harm to Time. Therefore, the balance of hardships weigh in Time's favor.

Conclusion

Time has established a trade dress in the *People* cover format which includes the condensed white lettering, and the display of the logo with a contrasting colored border. Globe has not shown that there is no likelihood of confusion, thus Time is likely to succeed on the merits of their claim of trademark infringement. Further, Time has shown that they will suffer irreparable harm if the injunction is not granted and that the balance of hardships tips in their favor. Further, for the reasons set forth above, Time's motion for a preliminary injunction is granted.

It is so ordered.

Patent and Trademark Office
Board of Patent Appeals and Interferences

Ex parte Gray

No. 88-0437

Decided August 17, 1988, and January 17, 1989

Released March 14, 1989

PATENTS

1. Patentability/Validity — Obviousness — In general (§115.0901)

Patentability/Validity — Obviousness — Relevant prior art (§115.0903)

Patent and Trademark Office does not have facilities for examining and comparing applicants' claimed human nerve growth factor, which is product-by-process claim, with prior art, and thus applicants had burden of persuasion to make some comparison between materials in order to establish unexpected properties for claimed factor, and applicants, having failed to do so, cannot contend on appeal that any doubt as to difference between two materials should be resolved in favor of patentability, since obviousness does not require absolute predictability.

2. Patentability/Validity — Obviousness — Relevant prior art (§115.0903)

Applicant can be required to prove that prior art products do not necessarily or inher-

ently possess characteristics of claimed product, and thus applicants, on appeal of rejection in which issue is whether prior art factor is identical or patentably indistinct from that of material on appeal, have burden of showing that inherency is not involved.

3. Patentability/Validity — Obviousness — Evidence of (§115.0906)

Mere conclusory statements in publication item are no more probative of non-obviousness than such statements would be in applicant's specification, and, even if such unverified statements were to be considered as those of expert in art, such statements would be inadequate in view of lack of any factual supporting evidence.

4. Patentability/Validity — Obviousness — In general (§115.0901)

Mere purity of claimed compound does not render substance unobvious.

5. Patentability/Validity — Obviousness — Relevant prior art (§115.0903)

Patentability/Validity — Adequacy of disclosure (§115.12)

Applicants whose claims for human nerve growth factor synthesized through use of recombinant DNA technology were rejected for obviousness must, in order to raise question of non-enablement of prior art, provide at very least declaration by person having ordinary skill in subject art that no method was known to that person prior to claimed invention whereby claimed material might have been synthesized.

Appeal from rejection of claims (Howard E. Schain, primary examiner, G.D. Draper, examiner).

Application for patent filed by Alane M. Gray and Axel Ullrich, serial no. 471,962, on March 3, 1983. From examiner's decision rejecting claims, applicants appeal. Affirmed.

Max D. Hensley, San Francisco, Calif., for appellants.

Before Pellman, Winters, and W. Smith, examiners-in-chief.

Pellman, examiner-in-chief.

This is an appeal from the examiner's decision finally rejecting claims 1, 11, 12, 13 and 15 through 18, remaining claims 19 and 20 having been withdrawn from consideration by the examiner. However, since, by amendment, claims 13, 15 and 16 were canceled, the claims before us for consideration are 1, 11, 12, 17 and 18.

The subject matter on appeal involves the human nerve growth factor β-NGF, identified by the particular amino acid sequence and being free from other proteins of human origin (claim 1). The invention also includes pharmaceutical compositions containing said nerve factor (claims 11 and 12) and said human nerve factor in which the amino acid sequence is preceded by a methionyl group (claims 17), as well as a composition containing the factor of claim 17 (claim 18). The particular human nerve factor of the present invention has been synthesized through the use of recombinant DNA technology and thus, is free from human proteins that would otherwise be expected to contaminate the composition. To describe the invention in greater detail and illustrate the claims on appeal, a copy of claim 1 is appended to this decision.

For evidence of obviousness, the references identified below are cited by the examiner.

Goldstein et al. (Goldstein) "Isolation of Human Nerve Growth Factor From Placental Tissue," Neurochemical Research *3*, 175-183 (1978)

Walker et al. (Walker), "Human Nerve Growth Factor: Lack of Immunocrossreactivity with Mouse Nerve Growth Factor," Life Sciences *26*, 195-200 (1980)

All of the claims stand rejected for being unpatentable (35 U.S.C. 103) in view of Goldstein or Walker. The examiner, at page 2 of the answer states that:

"Each of these prior art discloses human β-NGF that appears to be the same as that claimed wherein such was isolated from human placental tissue, versus the claimed β-NGF that was produced by recombinant techniques."

In the sentence bridging pages 2 and 3 of the answer, the examiner notes that the sequencing of a protein does not make the protein different, but merely constitutes a further characterization of the known material.

In response to the examiner's arguments, beginning at page 5 of the brief appellants

set forth their own arguments. Appellants apparently divide their argument into three points. The first is that the human β-NGF of the references is not inherently that of appellants. Second, appellants contend that the reference human nerve growth factor is not free from other human proteins. Finally, we are told that the cited prior art does not teach or suggest methionyl N-terminal human β-NGF.

At page 6 of their brief, appellants refer to the publication edited by Black, *Cellular and Molecular Biology of Neuronal Development,* Plenum Press, New York, Chapter 20, Breakefield et al, pages 309-328 (1984). Appellants refer specifically to page 310, wherein the author states that:

> "To establish whether patients with dysautonomia make an altered form of β-NGF, it is necessary to characterize the human form of this protein. This has been difficult, and although there are a number of reports on preliminary identification of a human NGF-like molecule (Goldstein *et al.*, 1978; Walker *et. al.*, 1980), no one has conclusively demonstrated its presence."

Appellants rely upon the foregoing as evidence that the Goldstein and Walker reports are merely preliminary and inconclusive.

At page 8 of their brief, appellants focus on the difference in the interpretation of Walker by the examiner vis-a-vis that by appellants as to the lack of immunological cross-reactivity between the art human β-NGF and murine β-NGF. Appellants conclude that the point is not whose theory is right, but contend that "when legitimate disputes arise about polypeptide identity they must-in view of prior board precedent, be resolved in appellants' favor. Rejections cannot be properly maintained on "maybe" references."

Beginning at page 9 of their brief, appellants raise the question of the purity of the claimed human β-NGF as compared with that of the references. Appellants explain that, due to the method of preparation, their nerve growth factor is free from any other human proteins. At page 10 of the brief, appellants query "why would the art be motivated to attempt to further purify human "β-NGF" beyond the level reported by Goldstein *et al* and Walker *et al*? If so motivated, does the art reasonably teach one of ordinary skill how to do so?"

In connection with the foregoing, at page 10 of the brief, appellants suggest that even if art were applied showing recombinant methods for synthesizing proteins, it would be clear from their discussion that more than a conventional recombinant method was involved in the preparation of the claimed human nerve growth factor.

At page 11 of their brief, appellants discuss the methionyl N-terminal human nerve growth factor. We are informed that "no reference of record teaches any reason for wanting to make methionyl β-NGF, and no reference teaches how to do so even if that was an objective. With regard to the therapeutic formulation of claim 18, there is no teaching or suggestion as to what sort of biological activity to reasonably expect from the methionyl N-terminal variant."

Although due consideration has been given to the opposing arguments and supporting evidence of appellants and of the examiner, we are unpersuaded of reversible error in the examiner's rejection, which will be sustained.

While the present claims are drafted in the form of a compound or a composition, the rationale underlying appellants' arguments is founded on the proposition that the claims are directed to a product-by-process. In any event, we are convinced that the legal philosophy employed in rejections involving products-by-process should be employed with respect to the claims before us. That is, insofar as we can observe, the difference between the material of Goldstein and of Walker and that claimed by appellants herein resides in the method of obtaining the human growth factor. The prior art material is recovered from natural sources and purified, while appellants' is produced by recombinant DNA methodology. However, the dispositive issue before us is whether the claimed factor exhibits any unexpected properties compared with that described by the cited publication items.

To answer the foregoing question, we turn to the decision in *In re Brown,* 59 CCPA 1036, 459 F.2d 531, 173 USPQ 685 (1972) wherein, at 59 CCPA 1041, Judge Baldwin, delivering the court's opinion, explains:

> "We are therefore of the opinion that when the prior art discloses a product which reasonably appears to be either identical with or only slightly different than a product claimed in a product-by-process claim, a rejection based alternatively on either section 102 or section 103 of the statute is eminently fair and acceptable. As a practical matter, the Patent Office is not equipped to manufacture

products by the myriad of processes put before it and then obtain prior art products and make physical comparisons therewith."

[1] Consistent with the court's holding, we find that, in the present case, the Office does not have the facilities for examining and comparing appellants' growth factor with that disclosed by Walker and by Goldstein. It is therefore entirely proper that appellants should have shouldered their burden of persuasion and made some comparison between the two materials to establish unexpected properties for the claimed factor. Having failed to do so, appellants are in a poor position now to contend that any doubt as to the difference between the two materials should be resolved in *favor* of patentability. Appellants do not inform us of the legal basis for their conclusion that this Board has held that doubt should be resolved in favor of an applicant and we are aware of no such recent decision. On the other hand, our reviewing tribunal, the United States Court of Customs and Patent Appeals in *In re Mixon,* 59 CCPA 1996, 470 F.2d 1374, 176 USPQ 296 (1973), responsive to Chief Judge Worley's discussion of the "rule of doubt", Judges Rich, Almond, Baldwin and Lane, in their concurring opinion state:

> "Since we have not been following any 'rule of doubt' policy and since that question is not involved in the present case we do not agree with the additional comments of the author."

In fact, rather than resolving doubt in favor of the applicant, the court has often held that obviousness does not require absolute predictability. See the decisions in *In re Merck and Company, Inc.,* 800 F.2d 1091, 231 USPQ 375 (Fed. Cir. 1986) and *In re Lamberti,* 545 F.2d 747, 192 USPQ 278 (CCPA 1976).

[2] At page 5 of their brief, appellants contend that human β-NGF, as described by Walker or Goldstein, is not "inherently" that of appellant. However, this has not been established. Appellants' attention is invited to the decision in *In re Best,* 562 F.2d 1252, 195 USPQ 430 (CCPA 1977), wherein the court held that the PTO can require an applicant to prove that the prior art products do not necessarily or inherently possess the characteristics of his claimed product. Accordingly, since the issue in the present appeal is whether the prior art factor is identical or patentably indistinct from that of the material on appeal, appellants have the burden of showing that inherency is not involved.

[3] Following the court's guidance in *In re Johnson,* 747 F.2d 1456, 223 USPQ 1260 (Fed. Cir. 1984), we have weighed the examiner's evidence of obviousness against appellants' countervailing evidence to determine whether the claims are patentable notwithstanding the references of record. It appears that appellants rely heavily upon the publication item by Breakefield, particularly pages 310 and 311. We are well aware that the author dismisses the "preliminary identification of a human β-NGF-like molecule" by Goldstein and Walker and concludes that no one has conclusively demonstrated its presence. Nevertheless, we do not interpret the subjective statement by Breakefield as adequate to overcome the specific findings reported by Goldstein and Walker. Mere conclusory statements in a publication item are no more probative of nonobviousness than would be said statements in appellants' specification. Compare *In re D'Ancicco,* 59 CCPA 748, 452 F.2d 1060, 172 USPQ 241 (1972). Moreover, even if we were to consider the unverified statements in the publication article as those of an expert in the art, the statements would be inadequate because of the lack of factual supporting evidence. Compare *In re Grunwell,* 609 F.2d 486, 203 USPQ 1055 (CCPA 1979).

The present situation is somewhat similar to that confronting the court in *Scripps Clinic & Research Foundation v. Genentech Inc.,* ——— F.Supp. ——— , 3 USPQ2d 1481 (DC NCalif 1987) wherein human blood-clotting factor VIII:C was involved. At page 3 USPQ2d 1487, the court points out that:

> Scripps also alleges infringement of product claims 24 through 29 covering concentrated preparations of "human Factor VIII:C.
>
> According to Scripps, these claims cover preparations, in the specified ranges of purity and potency, of Factor VIII:C with the functional and structural characteristics of the protein as it occurs naturally in humans.[6] Genentech, on the other hand, argues that Scripps' claims are limited to Factor VIII:C derived from human blood plasma. The issue posed is whether the asserted product claims must be interpreted to apply solely to concentrates of Factor VIII:C derived directly from human blood plasma or whether they extend also to other concentrates of Factor VIII:C having the same characteristics as those derived from human blood plasma.

In the paragraph bridging pages 1488–1489, the court explains that:

"the excerpts quoted by Genentech from deposition testimony of Drs. Katzmann and Zimmerman, Scripps' experts, that 'human' means 'obtained from human blood', are not probative on the issue of interpreting the claims. Dr. Katzmann's answer related to Factor V, not Factor VIII, and Dr. Zimmerman's answer did not purport to give an interpretation of the particular claim language. Human factor VIII:C as claimed in the patent therefore applies to any Factor VIII:C preparation, regardless of how produced, having the same material structural and functional characteristics as the plasma-derived preparation."

We are convinced that our decision herein is completely consistent with and supported by the above-noted holding of the District court in Northern California.

In the interest of completeness, we call attention to two other decisions that appear relevant to our present holding. The first of these is *In re Bergstrom,*[1] 57 CCPA 1240, 427 F.2d 1394, 166 USPQ 256 (1970), involving a rejection of certain pure prostaglandin compounds for not being novel in light of the material from which it was extracted. At page 57 CCPA 1250, the court held as follows:

"We need not decide the merits of that matter, for the fundamental error in the board's position, as we see it, is the analysis and answer it gave to the sole issue it accurately posed — 'whether the *claimed* pure materials are *novel* as compared with the *less pure* materials of the reference.' [emphasis supplied.] It seems to us that the answer to that question is self-evident: by definition, pure materials necessarily differ are the only ones existing and available as a standard of reference, as seems to be the situation here, perforce the 'pure' materials are 'new' with respect to them."

The other decision relevant to the facts before us, is In re *Wakefield,* 57 CCPA 959, 422 F.2d 897, 164 USPQ 636 (1970). In the *Wakefield* case, the claimed subject matter was synthetic rubber, while the prior art showed the corresponding naturally occurring product. At page 57 CCPA 966, Judge Lane, speaking for the court, disagreed with the board that the word "synthetic" as used in the claims would be applicable to purified natural product. In delivering the court's opinion, Judge Lane held that:

"we now turn to the examiner's view adopted by the Board, that the synthetic product is so similar to the natural product, purified to the extent allegedly shown in Davis, as to be '*prima facie* obvious'. We would agree with this conclusion as a tentative one based on similarity of structure and gross characteristics. However, such tentative conclusions of obviousness are rebutted in those instances where there was, at the time the invention was made, no known or obvious method of making the claimed composition, or where the claimed composition is found to possess unexpected characteristics. At least the first situation is present in the case before us, since it cannot be said that a method of making the claimed synthetic product would be known or obvious from Davis."

Although we acknowledge that our holding in the present case appears to be in conflict with the court's limited holding in the Wakefield appeal, we are convinced that our decision is consonant with the overwhelming weight of current patent jurisprudence involving questions of the type posed by appellants. Moreover, we point out that no objective evidence has been provided establishing that no method was known to those skilled in this field whereby the claimed material might have been synthesized. Therefore, although we have weighed all of the evidence and legal authorities, both pro and con, concerning the patentability of the claims on appeal, we find that the evidence and the weight of legal authority compels an affirmance of the examiner's rejection.

With respect to claims 17 and 18, the mere presence of a single methionyl moiety in a sequence of over 100 amino acids would not have been expected to alter the properties of the compound in a significant respect, in the absence of evidence to the contrary. It is our view that a minor inactive substituent on an otherwise unpatentable complex compound will not necessarily impart patentability to said compound. Thus, since we find claims 17 and 18 to be directed to an unpatentable modification of the compound to which the remaining claims are directed, these claims are held to be properly rejected for the same reasons as claims 1, 11 and 12.

For the reasons expressed above and those set forth in the answer, the examiner's decision rejecting claims 1, 11, 12, 17 and 18 is affirmed.

37 CFR 1.136(a) does not apply to the times for taking any subsequent action in connection with this appeal.

AFFIRMED.

[1] Cited by the court in footnote 6 in the *Scripps Clinic* decision.

awarded in the Court's discretion between $250.00 and $10,000.00 for each musical composition infringed. Plaintiffs argue that this was a willful violation of the copyright laws since plaintiffs put defendants on notice of the infringement, and defendants have refused to purchase a license. Plaintiffs suggest that the Court award $1,500 per infringement in this case, based upon the fact that if the defendants would have purchased the proper license, they would have paid $5,072.50 to plaintiffs as of December 1985. Plaintiffs argue that an award of $1,500 per infringement would serve as a deterrent to further infringements by not making it more financially prudent to fail to pay the license fee and then have a lesser amount of damages awarded in a copyright infringement action.

After a complete review of the record, this Court is of the opinion that plaintiffs shall be awarded $1,000.00 for each infringement, for a total amount of $7,000.00. As stated earlier, the defendants shall be jointly and severally liable for said statutory damages.

Finally, plaintiffs have requested an award of costs and attorneys' fees for the prosecution of this action. 17 U.S.C. §505 clearly allows that the Court may allow the recovery of full costs by any party and the Court may also allow reasonable attorneys' fees to the prevailing party as costs. This court is of the opinion that plaintiffs are entitled to costs and reasonable attorneys' fees in this action. Before the Court makes a final determination as to actual amount of costs and attorneys' fees, the Court will order that the plaintiffs' counsel file an affidavit showing the time expended by counsel in this matter and the costs incurred in prosecuting this action.

The Court being so advised, IT IS NOW ORDERED HEREIN AS FOLLOWS:

1. Motion of plaintiffs for summary judgment is hereby GRANTED.

2. Plaintiffs shall have fifteen (15) days from the date of entry of this order to FILE an affidavit showing the time expended by counsel in this matter and the costs incurred. Defendants shall then have fifteen (15) days thereafter to FILE any type of response concerning the reasonableness of the requests for attorneys' fees.

3. This action is hereby DISMISSED and STRICKEN from the docket of this Court.

4. A separate Judgment and Order shall be entered in accordance with this Memorandum Opinion and Order.

Patent and Trademark Office
Board of Patent Appeals and Interferences

Ex parte Forman, et al.

Appeal No. 602-90

Decided February 5, 1986 and April 22, 1986

Released July 2, 1986

PATENTS

1. Specification — Sufficiency of disclosure (§62.7)

Impossibility of determining, from reading of specification and record, number of mutant strains of original S. typhi and hyperconjugant strains of genetically engineered hybrid that are originally formed in each experiment, nor what amount of time, effort, and level of skill is needed to isolate single strains which can then be cloned to yield useful vaccines, supports conclusion that practice of invention would require undue experimentation.

Particular Patents — Vaccine

Forman, Baron, and Kopecko, application No. 289,013, Oral Vaccine for Immunization Against Enteric Disease, rejection of claims 1-33 affirmed.

Appeal from Art Unit 127.

Application for patent of Samuel B. Forman, Louis S. Baron, and Dennis J. Kopecko, application, Serial No. 289,013, filed July 31, 1981, for Oral Vaccine for Immunization Against Enteric Disease. From decision rejecting claims 1-33, applicants appeal. Affirmed.

William G. Gapcynski, Werten F.W. Bellamy, Francis A. Cooch and John H. Raubitschek, all of Washington, D.C., for appellants.

Before Serota, Milestone, and Goldstein, Examiners-in-Chief.

Goldstein, Examiner-in-Chief.

This appeal is from the examiner's final rejection of claims 1 to 33. Claims 34 to 40 have been allowed. Illustrative rejected claim 1 and allowed claim 34 are reproduced below.

1. A living attenuated oral vaccine, for the immunization against enteric disease, comprising as the active component an effective dose of a genetic hybrid derivative of an attenuated galactose epimeraseless mutant strain of S. typhi and at least one non-typhoid protective antigen carried thereby,

or a mixture of said genetic hybrid derivatives.

34. A genetic hybrid derivative, having deposit accession number ATTC 31904, of an attenuated galactose epimeraseless mutant strain of S. typhi and a non-typhoid protective antigen carried thereby, wherein the non-typhoid protection antigen is the form I antigen of Shigella sonnei and the genetic hybrid derivative strain expresses both S. typhi and S. sonnei antigens.

No references have been relied on by the examiner on appeal. References incorporated into the disclosure of the present specification by appellants are:

Germanier 3,856,935 Dec. 24, 1972
Kopecko, D.J. et al. (Kopecko), Infect. Immun. 29:207-214, 1980.

Claims 1 to 33 have been finally rejected under the first paragraph of 35 U.S.C. 112 as being based on an insufficiently enabling disclosure. We shall affirm this rejection.

Appellants' claimed invention is a class of oral vaccines consisting of hybrid bacteria which have been genetically engineered (specification, page 6, ultimate line) to produce an immunizing effect against both the bacterial antigen (typhoid) and one other non-typhoid enteric disease. A commercially available Salmonella typhi strain is converted into streptomycin resistant mutants by mutagenesis (specification, page 9, penultimate paragraph). The antigen for the non-typhoid disease is disclosed as being introduced into the mutated S. typhi strain through a variety of techniques utilizing both plasmid-borne genetic determinants and bacterial chromosomes (specification, page 6, last sentence of first paragraph). There is, however, no discussion in any detail whatever as to anything but the plasmid-borne embodiment. This embodiment involves introduction of the plasmid-borne antigen of the non-typhoid disease from a variety of mutant strains of enteric bacteria, such as, for example, Shigella, through a technique known as transconjugation (see paragraph bridging pages 7 and 8 of the specification, for example).

The basis of the examiner's rejection appears to be generally two-fold. It appears to be the examiner's position that, in the absence of a deposit, one of ordinary skill in the relevant art would not be able to obtain the specific S. typhi mutant apparently essential to appellants' invention. See specification page 20, lines 11-14. It also appears to be the examiner's position that the hyperconjugation procedure is not sufficiently well known and straightforward so that one of ordinary skill in this art could predict the results with an adequate degree of certainty based on only the record before her to justify the allowance of generic claims as opposed to those claims based on available, deposited, specific strains. For the reasons which we shall elaborate below, we conclude that on the record before us the examiner's position is correct.

The ultimate question in each case of this type is whether or not the specification contains a sufficiently explicit disclosure to enable one having ordinary skill in the relevant field to practice the invention claimed therein without the exercise of undue experimentation. The requirement for enablement can be found expressly stated in the first paragraph of 35 U.S.C. 112, which requires that the disclosure of an invention be "in such full, clear, concise and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same ...." The "undue experimentation" proscription is, in effect, a gloss on the statute which has arisen from decisional law which requires that sufficient information be given in the application so that one of ordinary skill in the art can practice it without the necessity for undue experimentation: Fields v. Conover, 58 CCPA 1366, 443 F.2d 1386, 170 USPQ 276 (1971); In re Colianni, 561 F.2d 220, 195 USPQ 150 (CCPA 1977). Also see Ansul Co. v. Uniroyal, Inc., 448 F.2d 872, 169 USPQ 759 (2d. Cir. 1971), cert. denied, 404 U.S. 1018, 172 USPQ 257 (1972). The principle that the disclosure of a patent application must enable practice of the invention claimed without undue experimentation was well established long before the enactment of the 1952 Patent Act. See Standard Brands, Inc. v. National Grain Yeast Corp., 101 F.2d 814, 40 USPQ 318 (3rd Cir. 1939), and cases cited therein.

The determination of what constitutes undue experimentation in a given case requires the application of a standard of reasonableness, having due regard for the nature of the invention and the state of the art: Ansul Co. v. Uniroyal, Inc., supra. The test is not merely quantitative, since a considerable amount of experimentation is permissible, if it is merely routine, or if the specification in question provides a reasonable amount of guidance with respect to the direction in which the experimentation should proceed to enable the determination of how to practice a desired embodiment of the invention claimed. The factors to be considered have been summarized as the quantity of experimentation necessary, the amount of direction or guidance presented, the presence or absence of working examples, the nature of the invention, the state of the prior art, the relative skill of those in that art, the predictability or unpredictability of the art and the breadth of the claims. In re Rainer, 52 CCPA 1593, 347 F.2d 574, 146 USPQ 218 (1965); In re Colianni, supra.

The examiner in this case has relief on the published decision in Ex parte Jackson, 217 USPQ 804 (PTO Bd. App. 1982), as supporting her position. Appellants on the other hand have pointed out the distinction that the issue in the Jackson case was the undue experimentation involved in finding additional strains over and above the deposited cultures in nature, whereas here the relevant hybrid bacterial strains are produced by genetic engineering by processes which at least begin with available bacterial strains. Although we recognize the distinction noted by appellants, we find that there is insufficient evidence of record in the present case to suggest that the burden of experimentation on one of ordinary skill in the relevant art would be any less in this case.

[1] We recognize that the level of skill in the art of molecular biology is quite high. Nonetheless, we also recognize the correctness of the examiner's position that experiments in genetic engineering produce, at best, unpredictable results. From a reading of appellants' specification and the Kopecko article of record,* it is impossible to determine how many mutant strains of the original S. typhi and hyperconjugant strains of the genetically engineered hybrid are originally formed in each experiment and how much time and effort and what level of skill must be exercised to isolate the single strains which are then cloned to yield the useful vaccines. If there are any known clues to assist one of ordinary skill in predicting which of the myriad strains that are presumably produced would be useful, there is nothing on the present record to suggest what they might be. We note that appellants have acknowledged (Paper No. 6, top of page 12) that "the genetic techniques involved are very time consuming" and that it is anticipated that it "will take about one year to construct" most strains according to this invention. It is readily apparent from our above discussion of the relevant law that time is not the sole criterion. However, in the present case there is also a lack of guidance leading to predictability as discussed in the immediately preceding portion of this opinion. The art under consideration appears to be relatively undeveloped. There do not appear to be any apparently precisely reproducible working examples, and most of the claims are broad in scope. Consideration of all of these factors leads to the conclusion that practice of the rejected claims would require undue experimentation. In re Rainer, supra; In re Colianni, supra.

As we have noted above, there does not appear to be on the present record, even including the Kopecko article, a single detailed example which could be followed by another worker in another lab to obtain a single specific microorganism (vaccine) within appellants' claims, without recourse to the deposited strains recited in the allowed claims. Appellants have requested that we indicate what microorganism deposits and supporting documentation would be required to result in the allowance of broader claims (see Reply to the Supplemental Examiner's Answer, Paper No. 20). Whether or not it would be proper for us to make such a determination in any event, it should be clear from our above discussion of the present record that we find it inadequate to enable us to respond even speculatively to appellants' request.

The decision of the examiner is affirmed.

AFFIRMED

ON REQUEST FOR RECONSIDERATION

This is a request for reconsideration of our decision of February 5, 1986, wherein we affirmed the examiner's rejection of claims 1 to 33.

We have carefully considered the arguments advanced by appellants but find nothing therein to convince us that the decision was in error. We have carefully considered particularly the section of the specification bridging pages 8 and 9, to which appellants referred in the request for reconsideration.

We note that construction of the donor strain involved four consecutive mating or transconjugation procedures. This procedure broadly may be well known, but there is nothing in evidence to indicate how complex the results might be in this particular case. Selection was required after each consecutive procedure, and there is no indication as to the frequency of the desired transfer in a cell population from which one could judge how much trial and error was involved in the selection.

We believe the examiner's characterization and ours of the procedures disclosed as being unpredictable was correct on this record. Appellants have added nothing to the record to convince us to the contrary but merely repeated arguments similar to those already presented before the examiner. There is no evidence properly before us to establish that one of ordinary skill in this art has actually been able to produce additional species representative of any of the claims broader than those allowed

* The additional literature citation (Hayes) and the Kopecko declaration under 37 CFR 1.132, to which it was attached, were considered untimely filed by the examiner and thus are not part of the record considered by us.

by the examiner. We emphasize that the declaration of Dennis J. Kopecko, filed under 37 C.F.R. 1.132, was refused consideration by the examiner as being untimely. Thus, it is not before us, and it would be inappropriate for us even to speculate upon its potential probative value.

It should be emphasized that the absence of a working example was not in itself stated to cause the specification to be insufficiently enabling. It was mentioned merely as one of the factors set forth in In re Rainer, 52 CCPA 1593, 347 F.2d 574, 146 USPQ 218 (1965), and In re Colianni, 561 F.2d 220, 195 USPQ 150 (CCPA 1977). Since there appears to be no direct control over the various mating and transconjugation procedures set forth in the specification (see also those pages following page 9, where additional selection is indicated) it may, in fact, be impossible to present an exactly reproducible working example in cases of this type. If true, however, this fact would merely further emphasize the requirement for appropriate deposits of useful organisms to enable the reasonably ready practice of inventions of this type.

The request has been considered but is denied with respect to making any change in our decision.

DENIED

Patent and Trademark Office
Trademark Trial and Appeal Board

Kraft, Inc. v.
Country Club Food Industries, Inc.

Decided May 27, 1986

TRADEMARKS

1. Class of goods — Particular cases — Similar (§67.2073)

Identity and similarity — Words — Similar (§67.4117)

Use of "Philadelphia's Famous Cheesecake" for cheesecake, and "Philadelphia Brand," or "Philly," for cream cheese, is likely to cause confusion.

2. Class of goods — How determined — In general (§67.2031)

"Philadelphia Brand," is so well known, in view of its use and promotion for over 100 years in connection with cream cheese, that purchasers assume common source when such mark is applied to goods that may not be so closely related that purchasers would otherwise ordinarily assume that they came from same source.

Trademark cancellation No. 13,549, by Kraft, Inc., against Country Club Food Industries, Inc., registration No. 1,200,046, issued June 29, 1982. Petition granted.

Nims, Howes, Collison & Isner, New York, N.Y., for Kraft, Inc.

Caesar, Rivise, Bernstein & Cohen, Ltd., Philadelphia, Pa., for Country Club Food Industries, Inc.

Before Rice, Allen, and Cissel, Members.

Cissel, Member.

Country Club Food Industries, Inc. registered the mark shown below on the Supplemental Register for "cheesecake for consumption on or off the premises."[1] Kraft, Inc. brought this action to cancel the registration.[2] The grounds set forth in the Petition for Cancellation are that petitioner has prior use and incontestable registrations of "PHILADELPHIA BRAND", in this form,[3] and "PHILLY"[4] for cream cheese, and that registrant's mark, as applied to registrant's goods, so resembles petitioner's marks, as they are used with petitioner's goods, the confusion is likely. Petitioner also alleged that at the time of registrant's filing the application which matured into the registration which petitioner seeks to cancel, registrant was not entitled to register its mark because it was not in compliance with the Federal Food, Drug and Cosmetic Act, 21 U.S.C. Sections 301, 321-392. No proof of this allegation was offered, however, and this point was not argued, so we treat it as conceded by petitioner.

---

[1] Reg. No. 1,200,046 issued June 29, 1982, based on an application filed July 31, 1978, claiming first use on July 7, 1978.

[2] Petition for Cancellation filed September 29, 1982.

[3] Reg. No. 392, 212, now owned by petitioner, issued December 16, 1941, was republished under Section 12(c) of the 1946 Act. Affidavits under Section 8 and Section 15 were accepted and received. The registration has been renewed twice, most recently on December 16, 1981.

[4] Reg. No. 777,517, now owned by petitioner, issued September 22, 1964. Combined affidavit under Sections 8 and 15 accepted. Renewed on September 22, 1984.

of Law will be entered on the same date herewith.

ORDER AND JUDGMENT

In accordance with the Findings of Fact and Conclusions of Law entered on the same date herewith,

IT IS HEREBY ORDERED AND ADJUDGED, as follows:

1. The Nolan patent (No. 4,506,189), issued on March 19, 1985, is a valid patent.

2. By the manufacture, production, sale and distribution of its SAF-T-COTE fluorescent lamp, Trojan has infringed the Nolan patent.

3. By virtue of this infringement, Shat-R-Shield is entitled to injunctive relief. Trojan shall immediately cease and desist from the manufacture, production, sale and distribution of the SAF-T-COTE fluorescent lamp.

4. Trojan shall recall all the SAF-T-COTE fluorescent lamps sold to and still in the possession of its customers.

5. The Court having determined that Trojan's infringement was not willful and wanton, Shat-R-Shield is not entitled to treble damages.

6. Shat-R-Shield shall have no accounting for monetary damages.

7. The Court having found that this is not an exceptional case, Shat-R-Shield is not entitled to its attorney's fees.

8. All claims having been resolved as to all parties herein, this action is now DISMISSED and STRICKEN from the docket.

9. There being no just reason for delay, this is a FINAL and APPEALABLE Order and Judgment.

Court of Appeals, Federal Circuit

In re Wands

No. 87-1454

Decided September 30, 1988

PATENTS

1. Patentability/Validity — Adequacy of disclosure (§115.12)

Data disclosed in application for immunoassay method patent, which shows that applicants screened nine of 143 cell lines developed for production of antibody necessary to practice invention, stored remainder of said cell lines, and found that four out of nine cell lines screened produced antibody falling within limitation of claims, were erroneously interpreted by Board of Patent Appeals and Interferences as failing to meet disclosure requirements of 35 USC 112, since board's characterization of stored cell lines as "failures" demonstrating unreliability of applicants' methods was improper in view of fact that such unscreened cell lines prove nothing concerning probability of success of person skilled in art attempting to obtain requisite antibodies using applicants' methods.

2. Patentability/Validity — Adequacy of disclosure (§115.12)

Disclosure in application for immunoassay-method patent does not fail to meet enablement requirement of 35 USC 112 by requiring "undue experimentation," even though production of monoclonal antibodies necessary to practice invention first requires production and screening of numerous antibody producing cells or "hybridomas," since practitioners of art are prepared to screen negative hybridomas in order to find those that produce desired antibodies, since in monoclonal antibody art one "experiment" is not simply screening of one hybridoma but rather is entire attempt to make desired antibody, and since record indicates that amount of effort needed to obtain desired antibodies is not excessive, in view of applicants' success in each attempt to produce antibody that satisfied all claim limitations.

Appeal from decision of Patent and Trademark Office, Board of Patent Appeals and Interferences.

Application for patent of Jack R. Wands, Vincent R. Zurawski, Jr., and Hubert J. P. Schoemaker, serial number 188,735. From decision of Board of Patent Appeals and Interferences affirming rejection of application, applicants appeal. Reversed; Newman, J., concurring in part and dissenting in part in separate opinion.

Jorge A. Goldstein, of Saidman, Sterne, Kessler & Goldstein (Henry N. Wixon, with them on brief), Washington, D.C., for appellant.

John H. Raubitschek, associate solicitor (Joseph F. Nakamura and Fred E. McKelvey, with him on brief), PTO, for appellee.

Before Smith, Newman, and Bissell, circuit judges.

Smith, J.

This appeal is from the decision of the Patent and Trademark Office (PTO) Board of Patent Appeals and Interferences (board) affirming the rejection of all remaining claims in appellant's application for a patent, serial No. 188,735, entitled "Immunoassay Utilizing Monoclonal High Affinity IgM

Antibodies," which was filed September 19, 1980.[1] The rejection under 35 U.S.C. §112, first paragraph, is based on the grounds that appellant's written specification would not enable a person skilled in the art to make the monoclonal antibodies that are needed to practice the claimed invention without undue experimentation. We reverse.

I. *Issue*

The only issue on appeal is whether the board erred, as a matter of law, by sustaining the examiner's rejection for lack of enablement under 35 U.S.C. §112, first paragraph, of all remaining claims in appellants' patent application, serial No. 188,735.

II. *Background*

A. *The Art.*

The claimed invention involves immunoassay methods for the detection of hepatitis B surface antigen by using high-affinity monoclonal antibodies of the IgM isotype. *Antibodies* are a class of proteins (immunoglobulins) that help defend the body against invaders such as viruses and bacteria. An antibody has the potential to bind tightly to another molecule, which molecule is called an antigen. The body has the ability to make millions of different antibodies that bind to different antigens. However, it is only after exposure of an antigen that a complicated *immune response* leads to the production of antibodies against that antigen. For example, on the surface of hepatitis B virus particles there is a large protein called *hepatitis B surface antigen* (HBsAg). As its name implies, it is capable of serving as an antigen. During a hepatitis B infection (or when purified HBsAg is injected experimentally), the body begins to make antibodies that bind tightly and specifically to HBsAg. Such antibodies can be used as reagents for sensitive diagnostic tests (*e.g.*, to detect hepatitis B virus in blood and other tissues, a purpose of the claimed invention). A method for detecting or measuring antigens by using antibodies as reagents is called an *immunoassay*.

Normally, many different antibodies are produced against each antigen. One reason for this diversity is that different antibodies are produced that bind to different regions (determinants) of a large antigen molecule such as HBsAg. In addition, different antibodies may be produced that bind to the same determinant. These usually differ in the tightness with which they bind to the determinant. *Affinity* is a quantitative measure of the strength of antibody-antigen binding. Usually an antibody with a higher affinity for an antigen will be more useful for immunological diagnostic tests than one with a lower affinity. Another source of heterogeneity is that there are several immunoglobulin classes or *isotypes*. Immunoglobulin G (IgG) is the most common isotype in serum. Another isotype, immunoglobulin M (IgM), is prominent early in the immune response. IgM molecules are larger than IgG molecules, and have 10 antigen-binding sites instead of the 2 that are present in IgG. Most immunoassay methods use IgG, but the claimed invention uses only IgM antibodies.

For commercial applications there are many disadvantages to using antibodies from serum. Serum contains a complex mixture of antibodies against the antigen of interest within a much larger pool of antibodies directed at other antigens. There are available only in a limited supply that ends when the donor dies. The goal of monoclonal antibody technology is to produce an unlimited supply of a single purified antibody.

The blood cells that make antibodies are *lymphocytes*. Each lymphocyte makes only one kind of antibody. During an immune response, lymphocytes exposed to their particular antigen divide and mature. Each produces a *clone* of identical daughter cells, all of which secrete the same antibody. Clones of lymphocytes, all derived from a single lymphocyte, could provide a source of a single homogeneous antibody. However, lymphocytes do not survive for long outside of the body in cell culture.

Hybridoma technology provides a way to obtain large numbers of cells that all produce the same antibody. This method takes advantage of the properties of *myeloma* cells derived from a tumor of the immune system. The cancerous myeloma cells can divide indefinitely in vitro. They also have the potential ability to secrete antibodies. By appropriate experimental manipulations, a myeloma cell can be made to fuse with a lymphocyte to produce a single hybrid cell (hence, a hybridoma) that contains the genetic material of both cells. The hybridoma secretes the same antibody that was made by its parent lymphocyte, but acquires the capability of the myeloma cell to divide and grow indefinitely in cell culture. Antibodies produced by a clone of hybridoma cells (i.e., by hybridoma

[1] *In re Wands*, Appeal No. 673-76 (Bd. Pat. App. & Int. Dec. 30, 1986).

cells that are all progeny of a single cell) are called monoclonal antibodies.[2]

B. *The Claimed Invention.*

The claimed invention involves methods for the immunoassay of HBsAg by using high-affinity monoclonal IgM antibodies. Jack R. Wands and Vincent R. Zurawski, Jr., two of the three coinventors of the present application, disclosed methods for producing monoclonal antibodies against HBsAg in United States patent No. 4,271,145 (the '145 patent), entitled "Process for Producing Antibodies to Hepatitis Virus and Cell Lines Therefor," which patent issued on June 2, 1981. The '145 patent is incorporated by reference into the application on appeal. The specification of the '145 patent teaches a procedure for immunizing mice against HBsAg, and the use of lymphocytes from these mice to produce hybridomas that secrete monoclonal antibodies specific for HBsAg. The '145 patent discloses that this procedure yields both IgG and IgM antibodies with high-affinity binding to HBsAg. For the stated purpose of complying with the best mode requirement of 35 U.S.C. §112, first paragraph, a hybridoma cell line that secretes IgM antibodies against HBsAg (the 1F8 cell line) was deposited at the American Type Culture Collection, a recognized cell depository, and became available to the public when the '145 patent issued.

The application on appeal claims methods for immunoassay of HBsAg using monoclonal antibodies such as those described in the '145 patent. Most immunoassay methods have used monoclonal antibodies of the IgG isotype. IgM antibodies were disfavored in the prior art because of their sensitivity to reducing agents and their tendency to self-aggregate and precipitate. Appellants found that their monoclonal IgM antibodies could be used for immunoassay of HbsAg with unexpectedly high sensitivity and specificity. Claims 1, 3, 7, 8, 14, and 15 are drawn to methods for the immunoassay of HBsAg using high-affinity IgM monoclonal antibodies. Claims 19 and 25-27 are for chemically modified (*e.g.*, radioactively labeled) monoclonal IgM antibodies used in the assays. The broadest method claim reads:

> 1. An immunoassay method utilizing an antibody to assay for a substance comprising hepatitis B-surface antigen (HBsAg) determinants which comprises the steps of:
>
> contacting a test sample containing said substance comprising HBsAg determinants with said antibody; and
>
> determining the presence of said substance in said sample;
>
> wherein said antibody is a monoclonal high affinity IgM antibody having a binding affinity constant for said HBsAg determinants of at least 10^9 M^{-1}.

Certain claims were rejected under 35 U.S.C. §103; these rejections have not been appealed. Remaining claims 1, 3, 7, 8, 14, 15, 19, and 25-27 were rejected under 35 U.S.C. §112, first paragraph, on the grounds that the disclosure would not enable a person skilled in the art to make and use the invention without undue experimentation. The rejection is directed solely to whether the specification enables one skilled in the art to make the monoclonal antibodies that are needed to practice the invention. The position of the PTO is that data presented by Wands show that the production of high-affinity IgM anti-HBsAg antibodies is unpredictable and unreliable, so that it would require undue experimentation for one skilled in the art to make the antibodies.

III. *Analysis*

A. *Enablement by Deposit of Micro-organisms and Cell Lines.*

The first paragraph of 35 U.S.C. §112 requires that the specification of a patent must enable a person skilled in the art to make and use the claimed invention. "Patents * * * are written to enable those skilled in the art to practice the invention."[3] A patent need not disclose what is well known in the art.[4] Although we review underlying facts found by the board under a "clearly erroneous" standard,[5] we review enablement as a question of law.[6]

Where an invention depends on the use of living materials such as microorganisms or

[2] For a concise description of monoclonal antibodies and their use in immunoassay see *Hybritech, Inc. v. Monoclonal Antibodies, Inc.*, 802 F.2d 1367, 1368-71, 231 USPQ 81, 82-83 (Fed. Cir. 1986), *cert. denied*, 107 S.Ct. 1606 (1987).

[3] *W.L. Gore & Assocs., Inc. v. Garlock, Inc.*, 721 F.2d 1540, 1556, 220 USPQ 303, 315 (Fed. Cir. 1983), *cert. denied*, 469 U.S. 851 (1984).

[4] *Lindemann Maschinenfabrik GMBH v. American Hoist & Derrick Co.*, 730 F.2d 1452, 1463, 221 USPQ 481, 489 (Fed. Cir. 1984).

[5] *Coleman v. Dines*, 754 F.2d 353, 356, 224 USPQ 857, 859 (Fed. Cir. 1985).

[6] *Moleculon Research Corp. v. CBS, Inc.*, 793 F.2d 1261, 1268, 229 USPQ 805, 810 (Fed. Cir. 1986), *cert. denied*, 107 S.Ct. 875 (1987); *Raytheon Co. v. Roper Corp.*, 724 F.2d 951, 960 n.6, 220 USPQ 592, 599 n.6 (Fed. Cir. 1983), *cert. denied*, 469 U.S. 835 [225 USPQ 232] (1984).

cultured cells, it may be impossible to enable the public to make the invention (i.e., to obtain these living materials) solely by means of a written disclosure. One means that has been developed for complying with the enablement requirement is to deposit the living materials in cell depositories which will distribute samples to the public who wish to practice the invention after the patent issues.[7] Administrative guidelines and judicial decisions have clarified the conditions under which a deposit of organisms can satisfy the requirements of section 112.[8] A deposit has been held necessary for enablement where the starting materials (i.e., the living cells used to practice the invention, or cells from which the required cells can be produced) are not readily available to the public.[9] Even when starting materials are available, a deposit has been necessary where it would require undue experimentation to make the cells of the invention from the starting materials.[10]

In addition to satisfying the enablement requirement, deposit of organisms also can be used to establish the filing date of the application as the prima facie date of invention,[11] and to satisfy the requirement under 35 U.S.C. §114 that the PTO be guaranteed access to the invention during pendency of the application.[12] Although a deposit may serve these purposes, we recognized, in *In re Lundak*,[13] that these purposes, nevertheless, may be met in ways other than by making a deposit.

A deposit also may satisfy the best mode requirement of section 112, first paragraph, and it is for this reason that the 1F8 hybridoma was deposited in connection with the '145 patent and the current application. Wands does not challenge the statements by the examiner to the effect that, although the deposited 1F8 line enables the public to perform immunoassays with antibodies produced by that single hybridoma, the deposit does not enable the generic claims that are on appeal. The examiner rejected the claims on the grounds that the written disclosure was not enabling and that the deposit was inadequate. Since we hold that the written disclosure fully enables the claimed invention, we need not reach the question of the adequacy of deposits.

B. *Undue Experimentation.*

Although inventions involving microorganisms or other living cells often can be enabled by a deposit,[14] a deposit is not always necessary to satisfy the enablement requirement.[15] No deposit is necessary if the biological organisms can be obtained from readily available sources or derived from readily available starting materials through routine screening that does not require undue experimentation.[16] Whether the specification in an application involving living cells (here, hybridomas) is enabled without a deposit must be decided on the facts of the particular case.[17]

Appellants contend that their written specification fully enables the practice of

---

[7] *In re Argoudelis*, 434 F.2d 1390, 1392-93, 168 USPQ 99, 101-02 (CCPA 1970).

[8] *In re Lundak*, 773 F.2d 1216, 227 USPQ 90 (Fed. Cir. 1985); *Feldman v. Aunstrup*, 517 F.2d 1351, 186 USPQ 108 (CCPA 1975), *cert. denied*, 424 U.S. 912 [188 USPQ 720] (1976); Manual of Patent Examining Procedure (MPEP) 608.01 (p)(C) (5th ed. 1983, rev. 1987). *See generally* Hampar, *Patenting of Recombinant DNA Technology: The Deposit Requirement*, 67 J. Pat. Trademark Off. Soc'y 569 (1985).

[9] *In re Jackson*, 217 USPQ 804, 807-08 (Bd. App. 1982) (strains of a newly discovered species of bacteria isolated from nature); *Feldman*, 517 F.2d 1351, 186 USPQ 108 (uncommon fungus isolated from nature); *In re Argoudelis*, 434 F.2d at 1392, 168 USPQ at 102 (novel strain of antibiotic-producing microorganism isolated from nature); *In re Kropp*, 143 USPQ 148, 152 (Bd. App. 1959) (newly discovered microorganism isolated from soil).

[10] *Ex parte Forman*, 230 USPQ 546, 547 (Bd. Pat. App. & Int. 1986) (genetically engineered bacteria where the specification provided insufficient information about the amount of time and effort required); *In re Lundak*, 773 F.2d 1216, 227 USPQ 90 (unique cell line produced from another cell line by mutagenesis).

[11] *In re Lundak*, 773 F.2d at 1222, 227 USPQ at 95-96; *In re Feldman*, 517 F.2d at 1355, 186 USPQ at 113; *In re Argoudelis*, 434 F.2d at 1394-96, 168 USPQ at 103-04 (Baldwin, J. concurring).

[12] *In re Lundak*, 773 F.2d at 1222, 227 USPQ at 95-96; *In re Feldman*, 517 F.2d at 1354, 186 USPQ at 112.

[13] *In re Lundak*, 773 F.2d at 1222, 227 USPQ at 95-96.

[14] *In re Argoudelis*, 434 F.2d at 1393, 168 USPQ at 102.

[15] *Tabuchi v. Nubel*, 559 F.2d 1183, 194 USPQ 521 (CCPA 1977).

[16] *Id.* at 1186-87, 194 USPQ at 525; *Merck & Co. v. Chase Chem. Co.*, 273 F.Supp. 68, 77, 155 USPQ 139, 146 (D.N.J. 1967); *Guaranty Trust Co. v. Union Solvents Corp.*, 54 F.2d 400, 403-06, 12 USPQ 47, 50-53 (D. Del. 1931), *aff'd*, 61 F.2d 1041, 15 USPQ 237 (3d Cir. 1932), *cert. denied*, 288 U.S. 614 (1933); MPEP 608.01(p)(C) ("No problem exists when the microorganisms used are known and readily available to the public.").

[17] *In re Jackson*, 217 USPQ at 807; *see In re Metcalfe*, 410 F.2d 1378, 1382, 161 USPQ 789, 792 (CCPA 1969).

their claimed invention because the monoclonal antibodies needed to perform the immunoassays can be made from readily available starting materials using methods that are well known in the monoclonal antibody art. Wands states that application of these methods to make high-affinity IgM anti-HBsAg antibodies requires only routine screening, and that does not amount to undue experimentation. There is no challenge to their contention that the starting materials (i.e., mice, HBsAg antigen, and myeloma cells) are available to the public. The PTO concedes that the methods used to prepare hybridomas and to screen them for high-affinity IgM antibodies against HBsAg were either well known in the monoclonal antibody art or adequately disclosed in the '145 patent and in the current application. This is consistent with this court's recognition with respect to another patent application that methods for obtaining and screening monoclonal antibodies were well known in 1980.[18] The sole issue is whether, in this particular case, it would require undue experimentation to produce high-affinity IgM monoclonal antibodies.

Enablement is not precluded by the necessity for some experimentation such as routine screening.[19] However, experimentation needed to practice the invention must not be undue experimentation.[20] "the key word is 'undue,' not 'experimentation.' "[21]

> The determination of what constitutes undue experimentation in a given case requires the application of a standard of reasonableness, having due regard for the nature of the invention and the state of the art. Ansul Co. v. Uniroyal, Inc. [448 F.2d 872, 878-79; 169 USPQ 759, 762-63 (2d Cir. 1971), *cert. denied*, 404 U.S. 1018 [172 USPQ 257] (1972)]. The test is not merely quantitative, since a considerable amount of experimentation is permissible, if it is merely routine, or if the specification in question provides a reasonable amount of guidance with respect to the direction in which the experimentation should proceed * * *.[22]

The term "undue experimentation" does not appear in the statute, but it is well established that enablement requires that the specification teach those in the art to make and use the invention without undue experimentation.[23] Whether undue experimentation is needed is not a single, simple factual determination, but rather is a conclusion reached by weighing many factual considerations. The board concluded that undue experimentation would be needed to practice the invention on the basis of experimental data presented by Wands. These data are not in dispute. However, Wands and the board disagree strongly on the conclusion that should be drawn from that data.

Factors to be considered in determining whether a disclosure would require undue experimentation have been summarized by the board in *Ex parte Forman*.[24] They include (1) the quantity of experimentation necessary, (2) the amount of direction or guidance presented, (3) the presence or absence of working examples, (4) the nature of the invention, (5) the state of the prior art, (6) the relative skill of those in the art, (7) the predictability or unpredictability of the art, and (8) the breadth of the claims.[25]

In order to understand whether the rejection was proper, it is necessary to discuss further the methods for making specific monoclonal antibodies. The first step for making monoclonal antibodies is to immunize an animal. The '145 patent provides a detailed description of procedures for immunizing a specific strain of mice against HBsAg. Next the spleen, an organ rich in lymphocytes, is removed and the lymphocytes are separated from the other spleen cells. The lymphocytes are mixed with myeloma cells, and the mixture is treated to cause a few of the cells to fuse with each other. Hybridoma cells that secrete the desired antibodies then must be isolated from the enormous number of other cells in the mixture. This is done through a series of screening procedures.

The first step is to separate the hybridoma cells from unfused lymphocytes and myeloma cells. The cells are cultured in a medi-

---

[18] *Hybritech*, 802 F.2d at 1384, 231 USPQ at 94.

[19] *Id.*; *Atlas Powder Co. v. E.I. DuPont De Nemours & Co.*, 750 F.2d 1569, 1576, 224 USPQ 409, 413 (Fed. Cir. 1984); *In re Angstadt*, 537 F.2d at 502-504, 190 USPQ at 218; *In re Geerdes*, 491 F.2d 1260, 1265, 180 USPQ 789, 793 (CCPA 1974); *Mineral Separation, Ltd. v. Hyde*, 242 U.S. 261, 270-71 (1916).

[20] *Hybritech*, 802 F.2d at 1384, 231 USPQ at 94; *W.L. Gore*, 721 F.2d at 1557, 220 USPQ at 316; *In re Colianni*, 561 F.2d 220, 224, 195 USPQ 150, 153 (CCPA 1977) (Miller, J., concurring).

[21] *In re Angstadt*, 537 F.2d at 504, 190 USPQ at 219.

[22] *In re Jackson*, 217 USPQ at 807.

[23] *See Hybritech*, 802 F.2d at 1384, 231 USPQ at 94; *Atlas Powder*, 750 F.2d at 1576, 224 USPQ at 413.

[24] *Ex parte Forman*, 230 USPQ at 547.

[25] *Id.*; *see In re Colianni*, 561 F.2d at 224, 195 USPQ at 153 (Miller, J., concurring); *In re Rainer*, 347 F.2d 574, 577, 146 USPQ 218, 221 (CCPA 1965).

um in which all the lymphocytes and myeloma cells die, and only the hybridoma cells survive. The next step is to isolate and clone hybridomas that make antibodies that bind to the antigen of interest. Single hybridoma cells are placed in separate chambers and are allowed to grow and divide. After there are enough cells in the clone to produce sufficient quantities of antibody to analyze, the antibody is assayed to determine whether it binds to the antigen. Generally, antibodies from many clones do not bind the antigen, and these clones are discarded. However, by screening enough clones (often hundreds at a time), hybridomas may be found that secrete antibodies against the antigen of interest.

Wands used a commercially available radioimmunoassay kit to screen clones for cells that produce antibodies directed against HBsAg. In this assay the amount of radioactivity bound gives some indication of the strength of the antibody-antigen binding, but does not yield a numerical affinity constant, which must be measured using the more laborious Scatchard analysis. In order to determine which anti-HBsAg antibodies satisfy all of the limitations of appellants' claims, the antibodies require further screening to select those which have an IgM isotype and have a binding affinity constant of at least 10^9 M^{-1}.[26] The PTO does not question that the screening techniques used by Wands were well known in the monoclonal antibody art.

During prosecution Wands submitted a declaration under 37 C.F.R §1.132 providing information about all of the hybridomas that appellants had produced before filing the patent application. The first four fusions were unsuccessful and produced no hybridomas. The next six fusion experiments all produced hybridomas that made antibodies specific for HBsAg. Antibodies that bound at least 10,000 cpm in the commercial radioimmunoassay were classified as "high binders." Using this criterion, 143 high-binding hybridomas were obtained. In the declaration, Wands stated that [27]

> It is generally accepted in the art that, among those antibodies which are binders with 50,000 cpm or higher, there is a very high likelihood that high affinity (Ka [greater than] 10^9 M^{-1}) antibodies will be found. However, high affinity antibodies can also be found among high binders of between 10,000 and 50,000, as is clearly demonstrated in the Table.

The PTO has not challenged this statement.

The declaration stated that a few of the high-binding monoclonal antibodies from two fusions were chosen for further screening. The remainder of the antibodies and the hybridomas that produced them were saved by freezing. Only nine antibodies were subjected to further analysis. Four (three from one fusion and one from another fusion) fell within the claims, that is, were IgM antibodies and had a binding affinity constant of at least 10^9 M^{-1}. Of the remaining five antibodies, three were found to be IgG, while the other two were IgM for which the affinity constants were not measured (although both showed binding well above 50,000 cpm).

Apparently none of the frozen cell lines received any further analysis. The declaration explains that after useful high-affinity IgM monoclonal antibodies to HBsAg had been found, it was considered unnecessary to return to the stored antibodies to screen for more IgMs. Wands says that the existence of the stored hybridomas was disclosed to the PTO to comply with the requirement under 37 C.F.R. §1.56 that applicants fully disclose all of their relevant data, and not just favorable results.[28] How these stored hybridomas are viewed is central to the positions of the parties.

The position of the board emphasizes the fact that since the stored cell lines were not completely tested, there is no proof that any of them are IgM antibodies with a binding affinity constant of at least 10^9 M^{-1}. Thus, only 4 out of 143 hybridomas, or 2.8 percent, were *proved* to fall within the claims. Furthermore, antibodies that were proved to be high-affinity IgM came from only 2 of 10 fusion experiments. These statistics are viewed by the board as evidence that appellants' methods were not predictable or reproducible. The board concludes that Wands' low rate of demonstrated success shows that a person skilled in the art would have to

---

[26] The examiner, the board, and Wands all point out that, technically, the strength of antibody-HBsAg binding is measured as *avidity*, which takes into account multiple determinants on the HBsAg molecule, rather than affinity. Nevertheless, despite this correction, all parties then continued to use the term "affinity." We will use the terminology of the parties. Following the usage of the parties, we will also use the term "high-affinity" as essentially synonymous with "having a binding affinity constant of at least 10^9 M^{-1}."

[27] A table in the declaration presented the binding data for antibodies from every cell line. Values ranged from 13,867 to 125,204 cpm, and a substantial proportion of the antibodies showed binding greater than 50,000 cpm. In confirmation of Dr. Wand's statement, two antibodies with binding less than 25,000 cpm were found to have affinity constants greater than 10^9 M^{-1}.

[28] *See Rohm & Haas Co. v. Crystal Chem. Co.*, 722 F.2d 1556, 220 USQ 98 (Fed. Cir. 1983).

engage in undue experimentation in order to make antibodies that fall within the claims.

Wands views the data quite differently. Only nine hybridomas were actually analyzed beyond the initial screening for HBsAg binding. Of these, four produced antibodies that fell within the claims, a respectable 44 percent rate of success. (Furthermore, since the two additional IgM antibodies for which the affinity constants were never measured showed binding in excess of 50,000 cpm, it is likely that these also fall within the claims.) Wands argues that the remaining 134 unanalyzed, stored cell lines should not be written off as failures. Instead, if anything, they represent partial success. Each of the stored hybridomas had been shown to produce a high-binding antibody specific for HBsAg. Many of these antibodies showed binding above 50,000 cpm and are thus highly likely to have a binding affinity constant of at least 10^9 M^{-1}. Extrapolating from the nine hybridomas that were screened for isotype (and from what is well known in the monoclonal antibody art about isotype frequency), it is reasonable to assume that the stored cells include some that produce IgM. Thus, if the 134 incompletely analyzed cell lines are considered at all, they provide some support (albeit without rigorous proof) to the view that hybridomas falling within the claims are not so rare that undue experimentation would be needed to make them.

The first four fusion attempts were failures, while high-binding antibodies were produced in the next six fusions. Appellants contend that the initial failures occurred because they had not yet learned to fuse cells successfully. Once they became skilled in the art, they invariably obtained numerous hybridomas that made high-binding antibodies against HBsAg and, in each fusion where they determined isotype and binding affinity they obtained hybridomas that fell within the claims.

Wands also submitted a second declaration under 37 C.F.R. §1.132 stating that after the patent application was submitted they performed an eleventh fusion experiment and obtained another hybridoma that made a high-affinity IgM anti-HBsAg antibody. No information was provided about the number of clones screened in that experiment. The board determined that, because there was no indication as to the number of hybridomas screened, this declaration had very little value. While we agree that it would have been preferable if Wands had included this information, the declaration does show that when appellants repeated their procedures they again obtained a hybridoma that produced an antibody that fit all of the limitations of their claims.

[1] We conclude that the board's interpretation of the data is erroneous. It is strained and unduly harsh to classify the stored cell lines (each of which was proved to make high-binding antibodies against HBsAg) as failures demonstrating that Wands' methods are unpredictable or unreliable.[9] At worst, they prove nothing at all about the probability of success, and merely show that appellants were prudent in not discarding cells that might someday prove useful. At best, they show that high-binding antibodies, the starting materials for IgM screening and Scatchard analysis, can be produced in large numbers. The PTO's position leads to the absurd conclusion that the more hybridomas an applicant makes and saves without testing the less predictable the applicant's results become. Furthermore, Wands' explanation that the first four attempts at cell fusion failed only because they had not yet learned to perform fusions properly is reasonable in view of the fact that the next six fusions were all successful. The record indicates that cell fusion is a technique that is well known to those of ordinary skill in the monoclonal antibody art, and there has been no claim that the fusion step should be more difficult or unreliable where the antigen is HBsAg than it would be for other antigens.

[2] When Wands' data is interpreted in a reasonable manner, analysis considering the factors enumerated in *Ex parte Forman* leads to the conclusion that undue experimentation would not be required to practice the invention. Wands' disclosure provides considerable direction and guidance on how to practice their invention and presents working examples. There was a high level of skill in the art at the time when the application was filed, and all of the methods needed to practice the invention were well known.

The nature of monoclonal antibody technology is that it involves screening hybridomas to determine which ones secrete antibody with desired characteristics. Practitioners of this art are prepared to screen negative hybridomas in order to find one that makes the desired antibody. No evidence was presented by either party on how many hybridomas would be viewed by those in the art as requiring undue experimentation to screen. However, it seems unlikely that un-

[9] Even if we were to accept the PTO's 2.8% success rate, we would not be required to reach a conclusion of undue experimentation. Such a determination must be made in view of the circumstances of each case and cannot be made solely by reference to a particular numerical cutoff.

due experimentation would be defined in terms of the number of hybridomas that were never screened. Furthermore, in the monoclonal antibody art it appears that an "experiment" is not simply the screening of a single hybridoma, but is rather the entire attempt to make a monoclonal antibody against a particular antigen. This process entails immunizing animals, fusing lymphocytes from the immunized animals with myeloma cells to make hybridomas, cloning the hybridomas, and screening the antibodies produced by the hybridomas for the desired characteristics. Wands carried out this entire procedure three times, and was successful each time in making at least one antibody that satisfied all of the claim limitations. Reasonably interpreted, Wands' record indicates that, in the production of high-affinity IgM antibodies against HBsAG, the amount of effort needed to obtain such antibodies is not excessive. Wands' evidence thus effectively rebuts the examiner's challenge to the enablement of their disclosure.[10]

IV. *Conclusion*

Considering all of the factors, we conclude that it would not require undue experimentation to obtain antibodies needed to practice the claimed invention. Accordingly, the rejection of Wands' claims for lack of enablement under 35 U.S.C. §112, first paragraph, is reversed.

REVERSED

Newman, J., concurring in part, dissenting in part.

A

I concur in the court's holding that additional samples of hybridoma cell lines that produce these high-affinity IgM monoclonal antibodies need not be deposited. This invention, as described by Wands, is not a selection of a few rare cells from many possible cells. To the contrary, Wands states that all monoclonally produced IgM antibodies to hepatitis B surface antigen have the desired high avidity and other favorable properties, and that all are readily preparable by now-standard techniques.

Wands states that his United States Patent No. 4,271,145 describes fully operable techniques, and is distinguished from his first four failed experiments that are referred to in the Rule 132 affidavit. Wands argues that these biotechnological mechanisms are relatively well understood and that the preparations can be routinely duplicated by those of skill in this art, as in *Hybritech, Inc. v. Monoclonal Antibodies, Inc.*, 802 F.2d 1367, 1380, 231 USPQ 81, 94 (Fed. Cir. 1986), *cert. denied*, 107 S.Ct. 1606 (1987). I agree that it is not necessary that there be a deposit of multiple exemplars of a cell system that is readily reproduced by known, specifically identified techniques.

B

I would affirm the board's holding that Wands has not complied with 35 U.S.C. §112, first paragraph, in that he has not provided data sufficient to support the breadth of his generic claims. Wands' claims on appeal include the following:

> 19. Monoclonal high affinity IgM antibodies immunoreactive with HBsAg determinants, wherein said antibodies are coupled to an insoluble solid phase, and wherein the binding affinity constant of said antibodies for said HBsAg determinants is at least 10^9 M^{-1}.
>
> 26. Monoclonal high affinity IgM antibodies immunoreactive with hepatitis B surface antigen.

Wands states that he obtained 143 "high binding monoclonal antibodies of the right specificity" in the successful fusions; although he does not state how they were determined to be high binding or of the right specificity, for Wands also states that only nine of these 143 were tested.

Of these nine, four (three from one fusion and one from another fusion) were found to have the claimed high affinity and to be of the IgM isotype. Wands states that the other five were either of a different isotype or their affinities were not determined. (This latter statement also appears to contradict his statement that all 143 were "high binding".)

Wands argues that a "success rate of four out of nine", or 44.4%, is sufficient to support claims to the entire class. The Commissioner deems the success rate to be four out of 143, or 2.8%; to which Wands responds with statistical analysis as to how unlikely it is that Wands selected the only four out of 143 that worked. Wands did not, however, prove the right point. The question is whether Wands, by testing nine out of 143 (the Commissioner points out that the randomness of the sample was not established), and finding that four out of the nine had the desired properties, has provided sufficient experimental support for the breadth of the requested claims, in the context that "experi-

[10] *In re Strahilevitz*, 668 F.2d 1229, 1232, 212 USPQ 561, 563 (CCPA 1982).

ments in genetic engineering produce, at best, unpredictable results", quoting from *Ex parte Forman*, 230 USPQ 546, 547 (Bd.Pat.App. and Int. 1986).

The premise of the patent system is that an inventor, having taught the world something it didn't know, is encouraged to make the product available for public and commercial benefit, by governmental grant of the right to exclude others from practice of that which the inventor has disclosed. The boundary defining the excludable subject matter must be carefully set: it must protect the inventor, so that commercial development is encouraged; but the claims must be commensurate with the inventor's contribution. Thus the specification and claims must meet the requirements of 35 U.S.C. §112. *In re Fisher*, 427 F.2d 833, 839, 166 USPQ 18, 23-24 (CCPA 1970).

As the science of biotechnology matures the need for special accommodation, such as the deposit of cell lines or microorganisms, may diminish; but there remains the body of law and practice on the need for sufficient disclosure, including experimental data when appropriate, that reasonably support the scope of the requested claims. That law relates to the sufficiency of the description of the claimed invention, and if not satisfied by deposit, must independently meet the requirements of Section 112.

Wands is not claiming a particular, specified IgM antibody. He is claiming all such monoclonal antibodies in assay for hepatitis B surface antigen, based on his teaching that such antibodies have uniformly reproducible high avidity, free of the known disadvantages of IgM antibodies such as tendency to precipitate or aggregate. It is incumbent upon Wands to provide reasonable support for the proposed breadth of his claims. I agree with the Commissioner that four exemplars shown to have the desired properties, out of the 143, do not provide adequate support.

Wands argues that the law should not be "harsher" where routine experiments take a long time. However, what Wands is requesting is that the law be less harsh. As illustrated in extensive precedent on the question of how much experimentation is "undue", each case must be determined on its own facts. *See, e.g., W.L. Gore & Assocs., Inc. v. Garlock, Inc.*, 721 F.2d 1540, 1557, 220 USPQ 303, 316 (Fed. Cir. 1983), *cert. denied*, 469 U.S. 851 (1984); *In re Angstadt*, 537 F.2d 498, 504, 190 USPQ 214, 218 (CCPA 1976); *In re Cook*, 439 F.2d 730, 734-35, 169 USPQ 298, 302-03 (CCPA 1971).

The various criteria to be considered in determining whether undue experimentation is required are discussed in, for example, *Fields v. Conover*, 443 F.2d 1386, 170 USPQ 276 (CCPA 1971); *In re Rainer*, 347 F.2d 574, 146 USPQ 218 (CCPA 1965); *Ex parte Forman*, 230 USPQ at 547. Wands must provide sufficient data or authority to show that his results are reasonably predictable within the scope of the claimed generic invention, based on experiment and/or scientific theory. In my view he has not met this burden.

Patent and Trademark Office
Trademark Trial and Appeal Board

In re Johanna Farms Inc.

Serial No. 542,343

Decided June 30, 1988

JUDICIAL PRACTICE AND PROCEDURE

1. Procedure — Prior adjudication — In general (§410.1501)

Trademark Trial and Appeal Board's prior decision upholding examiner's refusal to register proposed mark "La Yogurt" does not preclude registration of mark pursuant to subsequent application, since applicant, by presenting survey evidence and consumer letters regarding issue of how purchasers perceive proposed mark, has demonstrated that instant factual situation is different from situation presented in prior proceeding.

TRADEMARKS AND UNFAIR TRADE PRACTICES

2. Types of marks — Non-descriptive — Particular marks (§327.0505)

Term "La Yogurt," with "yogurt" disclaimed, is registrable, since word "yogurt" is common English generic term rather than corruption or misspelling of French word for yogurt, since examining attorney failed to meet burden of showing clear evidence of generic use of mark as whole, and since evidence of record, including survey and consumer letters to applicant, demonstrates that primary significance of "La Yogurt" to majority of relevant public is that of brand name rather than generic term.

lenges ETF's rights as to those goods but seeks to preclude a more expanded line of clothing or services being offered under the VITTORIO RICCI mark.

[3] Nina Ricci has shown that NINA RICCI has become an increasingly strong identification of source, that it has been vigilant in protecting its marks from encroachment by others, that NINA RICCI or related marks have been used on the goods and services at issue since well before ETF's limited use of its mark for such goods and services, that the goods of the parties would be sold in some of the same stores, and that overlap of the goods, which did not exist to any significant extent in prior years and does not, in fact, yet exist to any significant extent, would be fostered if ETF were allowed to expand into the field now occupied by Nina Ricci. Further, we see no basis for holding that Nina Ricci has estopped itself from challenging the right of expansion ETF asserts by reason of its registering NINA RICCI for shoes, a mark ETF did not oppose. We therefore conclude that the opposer, Nina Ricci, has met its burden of demonstrating that there exists a likelihood of confusion between its NINA RICCI and related marks and ETF's VITTORIO RICCI mark for the goods and services in question. In reaching this decision, we must reiterate the teaching of our predecessor court that there is "no excuse for even approaching the well-known trademark of a competitor ... and that all doubt as to whether confusion, mistake, or deception is likely is to be resolved against the newcomer, especially where the established mark is one which is famous. ..." *Planter's Nut & Chocolate Co. v. Crown Nut Co., Inc.*, 305 F.2d 916, 924-25, 134 USPQ 504, 511 (CCPA 1962). *See also Specialty Brands, Inc. v. Coffee Bean Distribs., Inc.*, 748 F.2d at 674, 223 USPQ at 1285. Based on the record before the Board, we must view ETF as the "newcomer" to the clothing and fashion accessories business despite the fact that ETF uses the established VITTORIO RICCI mark in the rather limited area of shoes and belts.

Accordingly, the decision of the Trademark Trial and Appeal Board is reversed.

REVERSED

U.S. Patent and Trademark Office
Board of Patent Appeals and Interferences

Ex parte Mark

No. 88-2811

Decided June 24, 1989

PATENTS

1. Patentability/Validity — Adequacy of disclosure (§115.12)

Claims, for cysteine-depleted muteins of biologically active proteins, which require mutein which is produced to retain biological activity of native protein, are enabling, in view of record establishing that, for given protein having cysteine residues, one skilled in art would be able to routinely determine whether deletion or replacement of cysteine residues would result in mutein which is within claims, and fact that given protein may not be amenable for use in present invention, in that cysteine residues are needed for biological activity of protein, does not militate against conclusion of enablement, since one skilled in art is clearly enabled to perform such work as needed to determine whether cysteine residues of any given protein are needed for retention of biological activity.

3. Practice and procedure in U.S. Patent and Trademark Office — Prosecution — Filing date (§110.0906)

Earliest filing date of parent applications to which present claims are entitled under 35 USC 120 is filing date of first parent application in chain which sets forth generic description of synthetic muteins of present invention.

Appeal from rejection of claims (Robin Teskin, examiner; Thomas Wiseman, supervisory patent examiner).

Applicants David F. Mark, Leo S. Lin, Shi-Da Yu Lu, and Alice M. Wang appeal from rejection of claims, application, serial no. 06/698,939 filed Feb. 7, 1985 (cysteine-depleted muteins of biologically active proteins). Reversed.

Albert P. Halluin and Jane R. McLaughlin, Emeryville, Calif., for appellants.

Before Goldstein, Pellman, and W. Smith, examiners-in-chief.

W. Smith, examiner-in-chief.

This is an appeal from the final rejection of claims 1 through 5 and 45 through 69. The appeal as to claims 54 and 56 was withdrawn

by appellants' counsel at oral argument.[1] Thus, claims 1 through 5, 45 through 53, 55 and 57 through 69 remain for our consideration, which are all of the claims remaining in the application.

Claims 1, 45, 54, 55, 56, 57 and 64 are illustrative of the subject matter on appeal and read as follows:

1. A synthetic mutein of a biologically active native protein in which native protein has at least one cysteine residue that is free to form a disulfide link and is nonessential to said biological activity, said mutein having at least one of said cysteine residues substituted by another amino acid and said mutein exhibiting the biological activity of said native protein.

45. A structural gene having a DNA sequence that encodes a synthetic mutein of a biologically active native protein which native protein has at least one cysteine residue that is free to form a disulfide link and is nonessential to said biological activity, said mutein having at least one of said cysteine residues substituted by another amino acid and said mutein exhibiting the biological activity of said native protein.

54. A method of preventing a protein having at least one cysteine residue that is free to form a disulfide link from forming said link comprising mutationally altering the protein by deleting the cysteine residue or replacing the cysteine residue with another amino acid.

55. The method of claim 54 wherein the protein is biologically active and the cysteine is not essential to said biological activity.

56. The method of claim 54 wherein the cysteine residue is replaced with serine or threonine.

57. A method for making a gene having a DNA sequence that encodes a synthetic mutein of a biologically active native protein which native protein has at least one cysteine residue that is free to form a disulfide link and is nonessential to said biological activity, said mutein having at least one of said cysteine residues substituted by another amino acid and said mutein exhibiting the biological activity of said native protein comprising:

(a) hybridizing single-stranded DNA comprising a strand of a structural gene that encodes said protein with a mutant oligonucleotide primer that is complimentary to a region of said strand that includes the codon for said cysteine residue or the antisense triplet paired with said codon, as the case may be, except for a mismatch with said codon or said antisense triplet which mismatch defines a triplet that codes for said other amino acid:

(b) extending the primer with DNA polymerase to form a mutational heteroduplex; and

(c) replicating said mutational heteroduplex.

64. An oligonucleotide for use in making a structural gene, said gene having a DNA sequence that encodes a synthetic mutein of a biologically active native protein which native protein has at least one cysteine residue that is free to form a disulfide link and is nonessential to said biological activity, and said mutein having at least one of said cysteine residues substituted by another amino acid and said mutein exhibiting the biological activity of said native protein, by oligonucleotide-directed mutagenesis, said oligonucleotide having a nucleotide sequence that is complimentary to a region of the strand of the structural gene that includes the codon for the cysteine residue or the anti-sense triple paired with said condon, as the case may be, except for a mismatch of said codon that defines a triplet that codes for said other amino acid.

The reference relied upon by the examiner is:

Mark et al.
(Mark '584) 4,518,584 May 21, 1985

A reference relied upon by the Board is:

Mark et al.
(Mark '585) 4,588,585 May 13, 1986

The sole rejection of the claims remaining on appeal is under 35 USC §112, first paragraph, as being nonenabled. In support of the rejection, the examiner relies upon a statement of prior art which appears at page 3, lines 22-line 34 of the present specification and at column 1, lines 55-56 of U.S. Patent No. 4,518,584 to Mark et al., one of the present parent applications, which reads as follows:

In this regard Shepard, H.M., et al, *Nature* (1981) *294*:563-565 describe a mutein of IFN-β in which the cysteine at position 141 of its amino acid sequence

[1] It became apparent at oral argument that appellants' invention revolves around the present synthetic muteins retaining the biological activity of the native protein. The method of claims 54 and 56 is not so limited. When this was brought to counsel's attention during oral argument, counsel orally withdrew claims 54 and 56 from appeal.

(there are three cysteines in native human IFN-β at positions 17, 31, and 141, *Gene* (1980) *10*:11-15 and *Nature* (1980) *285*:542-547) is replaced by tyrosine. This mutein was made by bacterial expression of a hybrid gene constructed from a partial IFN-β cDNA clone having a G ⊃ A transition at nucleotide 485 of the IFN-β gene. The mutein lacked the biological activity of native IFN-β leading the authors to conclude that the replaced cysteine was essential to activity.

In addition, the examiner relies upon a statement which appears in an amendment filed in co-pending, commonly assigned Serial No. 06/876,819 which reads as follows:

> The review of the newly allowed claims with the inventors in light of the presently available data concerning the claimed species revealed that the seven Cys to Ser substitutions possible within the mature CSF-1 sequence shown in Figure 5 each result in a substantial reduction in the *in vitro* colony stimulation assay specified in claim 53. Thus, the ser_{∞}CSF-1 species claimed in claim 20 (and also in claims 22 and 29) does not meet the requirement specified by claim 53. Nevertheless, applicants are of the view that the DNA encoding the ser_{∞}CSF-1 species as well as the other Cys substitution species may have other uses, as experimental probes for example. Accordingly, claim 20 which specifically claims ser_{∞}CSF-1 has been made independent. For the above described reasons, the ser_{∞}CSF-1 species has been deleted from claims 22 and 29.

The examiner's rejection is summarized at page 3 of the Examiner's Answer as follows:

> Essentially, the position taken in the rejection is that it would require undue further experimentation to construct by recombinant methods (site specific mutagenesis) the innumerable muteins encompassed by the instant claims (claims encompass modification of any protein which comprises a "non-essential" cysteine residue) and to screen the muteins produced for any of those which exhibit biological activity after modification.

The examiner further reasons that it is generally known in the art that cysteine residues facilitate the proper disulfide bonds and consequently the proper folding of a protein. The examiner concludes that it is likely that most of the muteins prepared by appellants' claimed methodology "would be inoperative simply because the removal of the cysteine would disturb proper folding of the molecule, thereby potentially blocking the active site or sites of the resulting mutein." (Examiner's Answer, page 4)

The examiner points out on page 6 of her Answer that the claims on appeal encompass any protein, even those which have not been characterized or cloned and that the mere sequencing of all possible proteins encompassed by the claims on appeal, would entail an undue amount of experimentation.

As set forth on page 7 of the Appeal Brief:

> Appellants' position is that given the disclosure of the present invention substituting a nonessential cysteine with a neutral amino acid, the nonessential cysteine residues of any candidate protein could be identified and substituted in ten days employing the methods disclosed in the instant disclosure and the general knowledge of the art at the time the application was filed. Such limited amount of experimentation based on the disclosure in the application and the success shown by three proteins certainly does not constitute undue experimentation.

These arguments are supported by the declaration of co-appellant Alice M. Wang filed under 37 CFR 1.132 on August 10, 1987. In her declaration Ms. Wang sets forth what she terms a reasonable scheme for determining which cysteine residues in a generic biologically active protein would be available for substitution without destroying the biological activity. The declaration sets forth a step-by-step scheme for implementing the claimed invention which parallels the disclosure of the present application.

The examiner sets forth on page 5 of her Answer that Ms. Wang's declaration does not refute the determination that undue experimentation is needed for implementation of the claimed invention because of the "limited successful embodiments shown and the established unpredictability associated with such modifications as to how many such site-specific mutageneses would need to be undergone to obtain even one alternative biologically active mutein."

[1] We have carefully considered the respective positions of the examiner and the appellants and find that we agree with appellants that the claims remaining on appeal are enabled by the present disclosure. The working examples of the present specification set forth experiments which establish that three proteins, IFN-β, IL-2 and TNF, have nonessential cysteine residues which may be deleted or replaced with the resulting mutein retaining the biological activity of the native protein. When it is considered that the claims remaining on appeal all require that the mutein produced retain the biological

activity of the native protein, we consider the disclosure of this application to be enabling. The passages relied upon by the examiner from Mark '584 and copending Serial No. 06/876,819 are merely examples of work which is *outside* the claims on appeal. The record before us establishes that for a *given* protein having cysteine residues, one skilled in the art would be able to routinely determine whether deletion or replacement of the cysteine residues would result in a mutein which is within the claims on appeal.

To the extent that the examiner is concerned that undue experimentation would be required to determine other proteins suitable for use in the present invention, we find Ms. Wang's declaration to be persuasive that only routine experimentation would be needed for one skilled in the art to practice the claimed invention for a given protein. The fact that a given protein may not be amenable for use in the present invention in that the cysteine residues are needed for the biological activity of the protein does not militate against a conclusion of enablement. One skilled in the art is clearly enabled to perform such work as needed to determine whether the cysteine residues of a given protein are needed for retention of biological activity.

The examiner's rejection under 35 USC §112 first paragraph, is reversed.

NEW GROUND OF REJECTION

Claims 1 through 5, 45 through 53, 55 and 57 through 67 are rejected under 35 USC §102(e) as being anticipated by Mark '584 or Mark '585.

The present application lists four co-inventors, Mark, Lin, Lu and Wang. Appellants state on page 1 of the present specification that this application has two lines of parent applications. It is the first line of parent applications, i.e., Serial Nos. 06/564,224, 06/486,162 and 06/435,154 which is of present interest.

This application is stated to be a continuation-in-part of Serial No. 06/564,224 which is a continuation-in-part of Serial No. 06/486,162 which is a continuation-in-part of Serial No. 06/435,154. Each of these parent applications lists only three inventors, Mark, Lin and Lu. Wang, who is a co-inventor of the present application, is not a co-inventor in the parent applications. Mark '584 issued from Serial No. 06/564,224. Mark '585 is a division of Serial No. 06/564,224, and shares common parentage with Mark '584 of Serial Nos. 06/486,162 and 06/435,154.

In order for the present claims to be entitled under 35 USC §120 to the benefit of the earlier filing date of any of the parent applications, their subject matter must be disclosed in the parent applications in the manner provided by 35 USC §112, first paragraph, including the description requirement of this section of the statute. *In re van Langenhoven*, 458 F.2d 132, 173 USPQ 426 (CCPA 1972).

[2] Here, our review leads us to the conclusion that the earliest filing date the present generic claims are entitled to is the December 20, 1983 filing date of parent application Serial No. 06/564,224 since this appears to be the first application in this chain which sets forth a generic description of the synthetic muteins of the present invention. Parent application Serial No. 06/486,162 describes only a synthetic mutein of IFN-β. The entire original disclosure of Serial No. 06/486,162 describes and is strictly limited to synthetic muteins of IFN-β except for original claim 20 of that application which was directed to "a nucleotide primer for mutagenesis, comprising an oligonucleotide of about 12 to about 24 bases." The specification of this application contains a corresponding disclosure of this generic nucleotide primer. However, comparing this generic disclosure of a nucleotide primer with that of the present application, i.e., claim 64, it is apparent that claim 20 of this parent application does not provide descriptive support for the *broader* oligonucleotide disclosed and claimed in this application. Thus, none of the present claims are entitled to the benefit of the earlier filing date of Serial No. 06/486,162, at best, only Serial No. 06/564,224.

Having made this determination, we find that Mark '584 or Mark '585 is available as prior art against the appealed claims under 35 USC §102(e) as these patents are by "others" having the effective filing date required by this section of the statute. The effective filing date of these two references, to the extent they disclose synthetic muteins of IFN-β, is October 19, 1982, the filing date of common parent application Serial No. 06/435,154. They are anticipatory of the claims included in this rejection in that these references describe the IFN-β synthetic mutein species of the present generic claims. *In re May*, 574 F.2d 1082, 1089, 197 USPQ 601, 607 (CCPA 1978).

Claims 68 and 69 are rejected under 35 USC §103 as being unpatentable over Mark '584 or Mark '585.

These claims are directed to a therapeutic formulation which comprises an effective amount of the mutein of the present inven-

tion and at least one other anti-cancer or anti-viral compound, e.g., gamma-interferon. While Mark '584 and Mark '585 describe such a therapeutic formulation, this description does not appear in common parent Serial No. 06/486,162. This parent application only indicates that the synthetic mutein of IFN-β is useful for treatment of viral infections, and various types of cancer where interferon therapy is indicated.

However, in view of the disclosed utility of the synthetic mutein of IFN-β as an anti-cancer or anti-viral compound in Serial No. 06/486,162, it would have been *prima facie* obvious to one of ordinary skill in the art to use the disclosed synthetic muteins of IFN-β in combination with other known anti-cancer or anti-viral compounds such as gamma-interferon. *In re Kerkhoven*, 626 F.2d 846, 205 USPQ 1069 (CCPA 1980).

Any request for reconsideration or modification of this decision by the Board of Patent Appeals and Interferences based upon the same record must be filed within one month from the date of the decision (37 CFR 1.197). Should appellants elect to have further prosecution before the examiner in response to the new rejection under 37 CFR 1.196(b) by way of amendment or showing of facts, or both, not previously of record, a shortened statutory period for making such response is hereby set to expire two months from the date of this decision.

37 CFR 1.136(a) does not apply to the times for taking any subsequent action in connection with this appeal.

REVERSED
37 CFR 1.196(b)

Court of Appeals, Federal Circuit

In re Iwahashi

No. 89-1019

Decided November 7, 1989

PATENTS

1. Patent grant — Inventions patentable (§105.05)

Assertion that claim includes or is directed to algorithm does not automatically require finding that claim is directed to non-statutory subject matter, since 35 USC 101 expressly includes processes as category of inventions which may be patented, and since every step-by-step process, whether it is electronic, chemical, or mechanical, involves algorithm in broad sense of term.

2. Patent grant — Inventions patentable (§105.05)

Fact that apparatus, in claim for auto-correlation circuit for use in pattern recognition, operates according to mathematical algorithm does not itself render claim non-statutory under 35 USC 101; rather, showing, when claim is analyzed as whole, that algorithm is implemented in specific manner to define structural relationships between claim's physical elements warrants finding that claim is directed to statutory subject matter.

Appeal from the U.S. Patent and Trademark Office, Board of Patent Appeals and Interferences.

Application for patent, serial no. 454,022, filed Dec. 28, 1982, by Hiroyuke Iwahashi, Yoshiki Nishioka, and Mitsuhiro Hakaridani, for auto-correlation circuit for use in pattern recognition. From decision rejecting claim, applicants appeal. Reversed.

Charles Gorenstein, of Birch, Stewart, Kolasch & Birch (Michael K. Mutter, with him on brief), Falls Church, Va., for appellant.

John C. Martin, associate solicitor (Fred E. McKelvey, solicitor, with him on brief), for appellee.

Before Nichols, senior circuit judge,* and Rich and Bissell, circuit judges.

Rich, J.

This appeal is from the decision of the United States Patent and Trademark Office (PTO) Board of Patent Appeals and Interferences (board), dated May 24, 1988, adhered to on reconsideration, affirming the examiner's final rejection of the single claim of applicants' patent application serial No. 454,022, filed December 28, 1982, entitled "Auto-Correlation Circuit for Use in Pattern Recognition." The sole ground of rejection is that the subject matter claimed is nonstatutory under 35 U.S.C. §101 because it is merely a mathematical algorithm. We reverse.

The real party in interest, according to appellants' brief, is Sharp Kabishiki Kaisha (Sharp Corporation).

---

* Judge Nichols heard oral argument but, due to illness, did not participate in the decision.

Court of Appeals, Federal Circuit

Atlas Powder Company v.
E.I. Du Pont De Nemours & Company

No. 84-504

Decided Dec. 27, 1984

PATENTS

1. Specification — Sufficiency of disclosure (§62.7)

Use of prophetic examples does not automatically make patent nonenabling, burden being on patent challenger to show by clear and convincing evidence that prophetic examples together with specification's other parts are nonenabling.

2. Infringement — Substitution of equivalents — Basic, improvement or paper patent (§39.753)

Where accused has appropriated material features of patent, infringement will be found even when those features have been supplemented and modified to such extent that accused may be entitled to patent for improvement.

3. Infringement — Substitution of equivalents — In general (§39.751)

Patentee that was unable effectively to use product that accused successfully developed, is not estopped from asserting infringement on equivalence theory, since focus in assessing equivalence is on whether accused's product performs substantially same as claimed product in function, way and result, and not on patentee's ability to devise product equivalent to patented product.

Particular patents — Explosives

3,447,978, Bluhm, Ammonium Nitrate Emulsion Blasting Agent and Method of Preparing Same, decision holding claims 1-5, 7, 12-14, and 16-17, valid and infringed, affirmed.

Appeal from District Court for the Northern District of Texas, Higgenbotham, J.; 221 USPQ 426.

Action by Atlas Powder Company, against E.I. Du Pont De Numours & Company, and Alamo Explosives Company, Inc., for patent infringement, in which defendant counterclaims for declaration of patent invalidity. From Judgment for plaintiff, defendants appeal. Affirmed.

Garland P. Andrews, Roy W. Hardin, David L. Hitchcock, and Richards, Harris & Medlock, all of Dallas, Tex., for plaintiff.

Stanley Neely, and Locke, Purnell, Boren, Laney & Neely, both of Dallas, Tex., and Lawrence F. Scinto, Nels T. Lippert, and Fitzpatrick, Cella, Harper & Scinto, all of New York, N.Y., for defendants.

Before Markey, Chief Judge, and Baldwin and Miller, Circuit Judges.

Baldwin, Circuit Judge.

This is an appeal by E. I. du Pont De Nemours & Co. and its customer Alamo Explosives Co., Inc. (collectively, "Du Pont"). The appeal is from a final judgment of the United States District Court for the Northern District of Texas holding product claims 1-5, 7, 12-14, and 16-17 of U.S. Patent No. 3,447,978 ('978 patent), issued to Harold Bluhm on June 3, 1969 and assigned to the Atlas Powder Co. ("Atlas"), not invalid under 35 U.S.C. §§ 102, 103, and 112, not fraudulently procured, and infringed. We affirm.

Background

The district court opinion, reported at 588 F.Supp. 1455, 221 USPQ 426 (1983), contains a detailed description of the facts, familiarity of which is presumed herein.

Briefly, the '978 patent relates to blasting agents, i.e., chemical mixtures that are relatively insensitive to normal modes of detonation but can be made to detonate with a high strength explosive primer. By the mid-1960's, blasting agents consisted of two major types: "ANFO" and "water-containing."

An "ANFO" blasting agent comprised a mixture of particulate ammonium nitrate, usually in the form of small round aggregates known as "prills," and fuel oil (e.g., diesel fuel). They were widely used in mining and construction because of their low cost, ease of handling, and ability to be mixed at the blast site rather than prepackaged at the plant. However, to work properly they could be used only in "dry" holes (without water) because water desensitized the mixutre, rendering it nondetonable.

A "water-containing" blasting agent, which was water resistant, generally comprised a slurry of particulate ammonium nitrate (or other oxidizing salt), a solid or liquid fuel, at least 5 percent water, and, as a sensitizer to increase explosive power, either a high explosive such as TNT or a chemical

such as nitric acid. Often, a gelling agent was added, particularly in the chemical sensitized slurries, to prevent the separation of sensitizers from slurry by forming a gel (a colloid in which the disperse phase has combined with the continuous phase to produce a viscous, jelly-like product). The use of sensitizers in water-containing blasting agents made preparation and handling more difficult and dangerous and, hence, more costly.

Before the '978 invention, Atlas manufactured a gelled slurry blasting agent called Aquanite, based on U.S. Patent No. 3,164,503, issued to Gehrig and assigned to Atlas. Aquanite used as a sensitizer nitric acid, which was highly caustic to skin and clothing and tended to separate out of the product even in the presence of a gelling agent, thereby reducing the product's stability and shelf life. Also, Aquanite was "hypergolic," i.e., it ignited wood, coal and various chemicals upon contact, which was suspected of causing the blasting agent to detonate prematurely.

The Invention

In 1965, Atlas assigned Harold Bluhm to investigate stabilizing its Aquanite gel. Bluhm experimented with various "emulsions" that did not contain nitric acid or a gelling agent. (An emulsion is a stable mixture of two immiscible liquids; a "water-in-oil" emulsion has a continuous oil and discontinuous aqueous phase; an "oil-in-water" emulsion is the reverse.) In early 1966, Bluhm formulated an intimately mixed water-in-oil, water resistant emulsion blasting agent. The product was sensitized with entrapped air rather than high explosives or chemicals and is the subject matter of the claims at issue. Representative is Claim 1:

1. An emulsion blasting agent consisting essentially of:

an aqueous solution of ammonium nitrate forming a discontinuous emulsion phase;

a carbonaceous fuel forming a continuous emulsion phase;

an occluded gas dispersed within said emulsion and comprising at least 4% by volume, thereof at 70°F. and atmospheric pressure; and

a water-in-oil type emulsifying agent;

said carbonaceous fuel having a consistency such that said occluded gas is held in said emulsion at a temperature of 70°F.

Claim 1 is the only independent claim in suit. The other, dependent claims describe various ingredients, such as microspheres for the occluded gas, additional fuels (e.g., aluminum), specific ranges of ingredients, and various properties of the blasting agent.

Du Pont's Activities

Du Pont sold a gelled slurry blasting agent until the latter part of the 1970's. In 1976, Du Pont formed a team to study the feasibility of an emulsion blasting agent. The team succeeded in making a water-in-oil emulsion blasting agent which Du Pont began making and selling in August 1978. Atlas sued for infringement in December 1979.

The District Court Proceedings

A non-jury trial was held between January 28 and February 2, 1982. Du Pont asserted invalidity of the '978 patent under sections 102(a), 103, and 112, "fraud" on the Patent and Trademark Office (PTO), and noninfringement. The district court rejected those assertions for the product claims at issue, holding that: (1) the claimed invention was not anticipated by the prior art; (2) the claimed invention would not have been obvious in view of the prior art; (3) the claims were not invalid for the patent's failure to comply with the "best mode," enablement, and "overclaiming" requirements of 35 U.S.C. §112; (4) the patent was not procured by "fraud" on the PTO; and (5) Du Pont's products infringed the claims under the doctrine of equivalence. On appeal, Du Pont contests those holdings, except for the one on best mode.

The district court denied Atlas increased damages and attorney fees because Du Pont had not willfully infringed the '978 patent claims and the case was not "exceptional." The district court also held that product claims 6, 13, and 15 were not infringed and that process claims 18-30 were invalid. Atlas has not appealed those holdings.

Issues

(1) Whether the district court was clearly erroneous in finding the invention of the patent claims at issue not anticipated by the prior art.

(2) Whether the district court erred in holding that the invention of the patent claims at issue would not have been obvious.

(3) Whether the district court erred in holding the patent claims at issue not invalid because of nonenablement.

(4) Whether the district court erred in holding no "fraud" on the PTO; i.e., no inequitable conduct.

(5) Whether the district court was clearly erroneous in finding that Du Pont's products infringed the '978 claims under the doctrine of equivalents.

Opinion

I. Standard of Review

The burden is on Du Pont, as appellant, to establish that the district court's ultimate fact findings (e.g., anticipation, infringement) were clearly erroneous, that the district court's legal conclusions (e.g., §103 obviousness, §112 enablement) were erroneous, or that the findings underlying the ultimate findings or conclusions were clearly erroneous. The "clearly erroneous" standard is satisfied if we are left with the firm conviction that error has been committed. See, e.g., Raytheon Co. v. Roper Corp., 724 F.2d 951, 956, 220 USPQ 592, 596 (Fed. Cir. 1983), cert. denied, 53 U.S.L.W. 3255 (U.S. Oct. 2, 1984).

II. Presumption of Validity

Under 35 U.S.C. §282, a patent is presumed valid, and the one attacking validity has the burden of proving invalidity by clear and convincing evidence. See, e.g., American Hoist & Derrick Co. v. Sowa & Sons, Inc., 725 F.2d 1350, 1360, 220 USPQ 763, 770 (Fed. Cir. 1984), cert. denied, 53 U.S.L.W. 3225 (U.S. Oct. 2, 1984). In that regard, the district court committed an error.

After correctly stating that the presumption of validity must be overcome with clear and convincing evidence, the district court stated that, if pertinent prior art were not cited to the PTO, as was the case here, the presumption is weakened and Du Pont must prove invalidity by only a preponderance of the evidence. That is incorrect. Though the introduction of prior art not before the PTO may facilitate meeting the challenger's ability to meet the burden of proof on invalidity, the presumption remains intact, the burden of persuasion remains on the challenger, and the "clear and convincing" standard does not change. See, e.g., Jervis B. Webb Co. v. Southern Systems, Inc., 742 F.2d 1388, 1392 & n.4, 222 USPQ 943, 945 & n.4 (Fed. Cir. 1984); Stratoflex, Inc. v. Aeroquip Corp., 713 F.2d 1530, 1534, 218 USPQ 871, 875 (Fed. Cir. 1983).

The error, however, was harmless. Indeed, it helped Du Pont at trial by lowering the standard of proof needed to prove its case. Even with the lower standard, Du Pont was unable to succeed.

III Anticipation

The district court's determination of no anticipation was a factual one that should be reversed only if appellant shows that it was clearly erroneous. See, e.g., Lindemann Maschinenfabrik GMBH v. American Hoist & Derrick Co., 730 F.2d 1452, 1458, 221 USPQ 481, 485 (Fed. Cir. 1984). Du Pont attempts to satisfy its burden by arguing that U.S. Patent No. 3,161,551, to Egly, et al., anticipated the claimed invention. We conclude, however, that the district court's finding of no anticipation was not clearly erroneous.

Egly, which Du Pont referred to at oral argument as the "closest prior art," describes an emulsion of ammonium nitrate, water, fuel oil, and water-in-oil emulsifying agent. Though Egly teaches the presence of solid ammonium nitrate prills as an essential ingredient, Du Pont argues that the '978 claims, because of the phrase "consisting essentially of," does not exclude the presence of those prills. See, e.g., In re Herz, 537 F.2d 549, 551, 190 USPQ 461, 463 (CCPA 1976); In re Janakirama-Rao, 317 F.2d 951, 954, 137 USPQ 893, 896 (CCPA 1963). Du Pont is correct. However, the district court found that Egly "does not mention air or gas as an ingredient in their explosives" and occluded air is an element of the claims. Hence, there is no anticipation under §102, because the exclusion of a claimed element from a prior art reference is enough to negate anticipation by that reference. Kalman v. Kimberly-Clark Corp., 713 F.2d 760, 771-72, 218 USPQ 781, 789 (Fed. Cir. 1983).

Du Pont asserts that Bluhm conceded in answer to an interrogatory that the first reduction to practice of the claimed invention was on January 14, 1966, and that Mr. Bluhm's notebook shows the composition prepared on that date to be identical to Egly's i.e., an emulsion without occluded air. Because the first reduction to practice was identical to Egly's product, Du Pont argues, the claimed invention is anticipated by Egly. Atlas argues that the notebook entry reveals that occluded air was present in the composition prepared on January 14, 1966, and hence, the first reduction to practice was not identical to Egly's composition. Atlas appears to be correct but, in any event, the district court's anticipation analysis properly focused on the claimed invention, which includes occluded air, not on Atlas' characterization of the January 14, 1966 experiment as the first reduction to practice.

IV. Obviousness

Though an invention is not anticipated by 35 U.S.C. §102, a patent should not issue if the differences between the claimed invention and prior art are such that the invention as a whole would have been obvious to one of ordinary skill in the art at the time the invention was made. 35 U.S.C. §103. In assessing nonobviousness a court should answer certain factual inquiries: (1) the scope and content of the prior art; (2) the level of ordinary skill in the art; (3) the differences between the claimed invention and the prior art; and (4) so-called "secondary" considerations, e.g., long felt need, unexpected results, commercial success. See, e.g., Stratoflex, Inc. v. Aeroquip Corp., 713 F.2d at 1538, 218 USPQ at 876; Simmons Fastener Corp. v. Illinois Tool Works, Inc., 739 F.2d 1573, 1575, 222 UPSQ 743, 746 (Fed. Cir. 1984). The "secondary" considerations, when present, may assist the court in determining nonobviousness without falling prey to hindsight reasoning.

Here, the district court made findings on the content of the prior art, the level of ordinary skill in the art, the differences between the prior art and the claimed invention as a whole, and then concluded that the claimed invention was nonobvious. Du Pont has not shown error in the legal conclusion of nonobviousness, or clear error in the underlying findings.

Content of the Prior Art and Differences Between It and the Claimed Invention

In addition to Egly, discussed above, the district court considered several patents and articles.

Atlas' Gehrig patent describes a blasting agent containing particulate ammonium nitrate, a solution of nitric acid in water, and fuel oil. Though the mixture may be an emulsion, the primary thrust of Gehrig is using a gel. Gehrig notes that, when an emulsion is used, the product quickly separates into its various components. Gehrig recommends that the emulsion be used within 24 hours to avoid separation. The gel form is considered desirable to stabilize the product for storage.

The claimed invention differs from Gehrig because Gehrig requires nitric acid as an essential ingredient. The '978 claims exclude the presence of nitric acid because the essence of the claimed composition is the elimination of nitric acid and the claim phrase "consisting essentially of" excludes ingredients that would "materially affect the basic and novel characteristics" of the claimed composition. In re Herz, 537 F.2d at 551, 190 USPQ at 463; In re Janakirama-Rao, 317 F.2d at 954, 137 USPQ at 895.

Gehrig does not teach substituting nitric acid with air to sensitize the product. Though it suggests the use of microballoons containing air as a stabilizer, it also discusses heating the product to remove entrapped air.

U.S. Patent No. 3,052,578, to Davis, describes a blasting agent comprising a blend of fuel oil and ammonium nitrate poured over solid ammonium nitrate. An oil-in-water, not water-in-oil, emulsifying agent is suggested to disperse the fuel. Though an emulsifying agent is used for dispersing purposes, the reference does not discuss forming an emulsion, and it does not suggest use of occluded air.

Two papers by Coxon relate to water resistant blasting agents. The first describes a water-in-oil emulsion of fuel oil and ammonium nitrate poured over solid ammonium nitrate. The second is similar, but prefers an oil-in-water emulsifying agent. Neither paper teaches the presence of occluded air; instead, the blasting agent requires solid ammonium nitrate. Thus, both Coxon papers, as well as Davis, are similar to Egly.

U.S. Patent No. 3,004,842, to Rowlinson, describes melting solid ammonium nitrate and mixing it with fuel oil and an emulsifying agent to form a solid blasting agent. A small amount of water may be added to reduce the melting point of the ammonium nitrate. Foaming agents can be added to increase the product's sensitivity.

U.S. Patent No. 3,453,158, to Clay, describes a gel or thickened slurry containing aqueous ammonium nitrate, a gelling agent or thickener, air bubbles serving as a sensitizer, and particulate fuels or sensitizers. The district court found that Clay does not use an emulsion, let alone a water-in-oil emulsion, and that finding has not been shown to be clearly erroneous.

Level of Skill in the Art

The district court found that the person of ordinary skill in the art would be one skilled in the art of explosives formulation, having knowledge of and experience with the chemical and physical properties of explosives. The person should be a chemist or chemical engineer with at least a bachelor's degree and several years of practical experience. Also, he or she should have a working knowledge of the principles of emulsion chemistry as applied to explosives formulation.

"Secondary" Considerations

The district court stated that, in light of "substantial differences" between the prior art and the product claims, it is not necessary to consider secondary factors, though they were raised by Atlas. Hence, the district court's opinion does not contain a section on "secondary criteria" or otherwise attempt to identify such criteria under the label. Nevertheless, the district court found that "[t]he Bluhm patent solved the problem of finding a water resistent ANFO blasting agent that did not require chemicpl sensitizers." Moreover, the district court in essence found that the solution to the problem was unexpected.

Though the prior art describes water-in-oil emulsions containing dissolved ammonium nitrate, fuel oil, and a water-in-oil emulsifying agent, the district court found that the art does not suggest that the emulsion itself can serve as a blasting agent. Egly, for example, teaches that such an emulsion -- without occluded air -- serves as a sensitizer that can be poured over solid ammonium nitrate to form a blasting agent. Gehrig teaches that the emulsion serves as a blasting agent only in the presence of nitric acid. That the Egly sensitizer itself serves as a blasting agent when occluded air is added, or that the Gehrig blasting agent could serve in that capacity without nitric acid, was unexpected. Though occluded air was recognized as an ingredient that could be included in blasting agent compositions, e.g., to stabilize the nitric acid containing product of Gehrig, the district court found that the references simply did not teach "that aeration can substitute for chemical sensitizers [e.g., nitric acid] in slurry explosives or that a water-in-oil emulsion is the most efficient system for entraining air."

Moreover, the district court found (and it has not been shown to be clearly erroneous) that the references cited by Du Pont deemphasize occluded air in emulsions and, hence, teach away from the importance of aeration. Egly and Davis do not mention air or gas as an ingredient in their explosives, and one of the Coxon papers teaches that detonation performance may be improved by using emulsifiers to eliminate frothing (air) from explosives.

Conclusion on Nonobviousness

In light of the differences between the claimed invention and prior art, the '978 solution to a troublesome problem, and the unexpected result that a water-in-oil emulsion of ammonium nitrate, fuel oil, and a water-in-oil emulsifying agent can serve as a blasting agent in the presence of occluded air, we agree with the district court's conclusion of nonobviousness.

Du Pont argues that it would have been obvious in 1966 to leave the nitric acid sensitizer out of Gehrig's slurry, intimately mix the fuel oil and ammonium nitrate, and sensitize the product in some other way, e.g., with air. We agree with the district court, however, that neither Gehrig nor the other prior art suggests those changes to obtain an emulsion blasting agent. As stated by the district court:

> It is quite a leap from recognition that dry ANFOs could be sensitized by aeration to realization that if an ANFO slurry was placed in the proper form of a water-in-oil emulsion and aerated, it would not require chemical sensitizers for detonability. This leap would not have been obvious in 1966.

V. Enablement

The district court rejected Du Pont's arguments of "overly broad," "overclaiming," and "non-enablement," and its argument that the broad scope of the claims is not supported by the limited disclosure present. In essence, those arguments are one: the '978 disclosure does not enable one of ordinary skill in the art to make and use the claimed invention, and hence, the claimed invention is invalid under 35 U.S.C. §112, ¶1.

To be enabling under §112, a patent must contain a description that enables one skilled in the art to make and use the claimed invention. Raytheon Co. v. Roper Corp., 724 F.2d at 960, 220 USPQ at 599. That some experimentation is necessary does not preclude enablement; the amount of experimentation, however, must not be unduly extensive. See, e.g., W.L. Gore & Associates, Inc. v. Garlock, Inc., 721 F.2d 1540, 1557, 220 USPQ 303, 316 (Fed. Cir. 1983), cert. denied, 53 U.S.L.W. 3226 (U.S. Oct. 2, 1984); In re Angstadt, 537 F.2d 498, 503, 190 USPQ 214, 218 (CCPA 1976). Determining enablement is a question of law. Raytheon Co. v. Roper Corp., 724 F.2d at 959-60, 220 USPQ at 599.

Du Pont argues that the patent disclosure lists numerous salts, fuels, and emulsifiers that could form thousands of emulsions but there is no commensurate teaching as to which combination would work. The disclosure, according to Du Pont, is nothing more than " a list of candidate ingredients" from which one skilled in the art would have to select and experiment unduly to find an operable emulsion.

The district court held it would have been impossible for Bluhm to list all operable emulsions and exclude the inoperable ones.

Further, it found such list unnecessary, because one skilled in the art would know how to select a salt and fuel and then apply "Bancroft's Rule" to determine the proper emulsifier. Bancroft's Rule was found by the district court to be a "basic principle of emulsion chemistry," and Du Pont has not shown that finding to be clearly erroneous.

We agree with the district court's conclusion on enablement. Even if some of the claimed combinations were inoperative, the claims are not necessarily invalid. "It is not a function of the claims to specifically exclude • • • possible inoperative substances • • • •" In re Dinh-Nguyen, 492 F.2d 856, 858-59, 181 USPQ 46, 48 (CCPA 1974) (emphasis omitted). Accord, In re Geerdes, 491 F.2d 1260, 1265, 180 USPQ 789, 793 (CCPA 1974); In re Anderson, 471 F.2d 1237, 1242, 176 USPQ 331, 334-35 (CCPA 1973). Of course, if the number of inoperative combinations becomes significant, and in effect forces one of ordinary skill in the art to experiment unduly in order to practice the claimed invention, the claims might indeed be invalid. See, e.g., In re Cook, 439 F.2d 730, 735, 169 USPQ 298, 302 (CCPA 1971). That, however, has not been shown to be the case here.

Du Pont contends that, because the '978 examples are "merely prophetic," they do not aid one skilled in the art in making the invention.[1] Because they are prophetic, argues Du Pont, there can be no guarantee that the examples would actually work.

[1] Use of prophetic examples, however, does not automatically make a patent nonenabling. The burden is on one challenging validity to show by clear and convincing evidence that the prophetic examples together with other parts of the specification are not enabling. Du Pont did not meet that burden here. To the contrary, the district court found that the "prophetic" examples of the specification were based on actual experiments that were slightly modified in the patent to reflect what the inventor believed to be optimum, and hence, they would be helpful in enabling someone to make the invention.

Du Pont argues that of some 300 experiments performed by Atlas before the filing of the '978 patent application, Atlas' records indicated that 40 percent failed "for some reason or another." The district court agreed that Atlas' records showed 40 percent "failed," but found that Atlas' listing of an experiment as a "failure" or "unsatisfactory" was misleading. Experiments were designated "failures," the district court found, in essence because they were not optimal under all conditions, but such optimality is not required for a valid patent. Decca Ltd. v. United States, 544 F.2d 1070, 1077, 191 USPQ 439, 444-45 (Ct. Cl. 1976). Accord, E. I. du Pont de Nemours & Co. v. Berkley & Co., 620 F.2d 1247, 1260, 205 USPQ 5, 10 (8th Cir. 1980). Cf. Raytheon Co. v. Roper Co., 724 F.2d at 958, 220 USPQ at 598. The district court also found that one skilled in the art would know how to modify slightly many of those "failures to form a better emulsion. Du Pont has not persuaded us that the district court was clearly erroneous in those findings.

Du Pont asserts that Atlas was able to produce suitable emulsions with only two emulsifiers, "Atmos 300" and "Span 80," and therefore, the disclosure should be construed to read upon only those two emulsifiers. However, Du Pont did not prove that the other disclosed emulsifiers were inoperable. The district court credited testimony by Atlas' expert, Dr. Fowkes, to the effect that he had successfully formed a number of detonable emulsions using a variety of emulsifiers specified in the '978 patent. Further, the district court found that one skilled in the art would know which emulsifiers would work in a given system. Indeed, the district court found that Du Pont's own researchers had little difficulty in making satisfactory emulsions with the emulsifying agents, salts, and fuels listed in the '978 patent. Those findings have not been shown to be clearly erroneous.

In sum, we conclude that Du Pont has failed to show that the district court erred in determining enablement.

VI. Inequitable Conduct

This court has held "inequitable conduct" in the PTO to be a more appropriate label than "fraud." J. P. Stevens & Co. v. Lex Tex Ltd., Nos. 84-754,-761, slip op. at 9, 223 USPQ 1089, 1092 (Fed. Cir. Nov. 9, 1984). Hence, this opinion will use the phrase "inequitable conduct" rather than "fraud."

Inequitable conduct requires proof by clear and convincing evidence of a threshold degree

[1] The PTO Manual of Patent Examining Procedure (MPEP) §608.01(p)(D) (5th ed. 1983), states

> Simulated or predicted test results and prophetical examples (paper examples) are permitted in patent applications. Working examples correspond to work actually performed and may describe tests which have actually been conducted and results that were achieved. Paper examples describe the manner and process of making an embodiment of the invention which has not actually been conducted. Paper examples should not be represented as work actually done. Paper examples should not be described using the past tense.

of materiality of the nondisclosed or false information. That threshold can be established by any of four tests: (1) objective "but for"; (2) subjective "but for"; (3) "but it may have been"; and (4) 37 C.F.R. §1.56(a), i.e., whether there is a substantial likelihood that a reasonable examiner would have considered the omitted or false information important in deciding whether to allow the application to issue as a patent. Slip op. at 10, 223 USPQ 1094–95. The PTO standard is the appropriate starting point because it is the broadest and most closely aligns with how one ought to conduct business with the PTO. Id.

Inequitable conduct also requires proof of a threshold intent. That intent need not be proven with direct evidence. It may be proven by showing acts the natural consequences of which are presumably intended by the actor. Id. Proof of deliberate scheming is not needed; gross negligence is sufficient. Gross negligence is present when the actor knew or should have known of the materiality of a withheld reference. Id. at 11, 223 USPQ at 1096. On the other hand, simple negligence, oversight or an erroneous judgment made in good faith is insufficient. Id.

Once the thresholds of materiality and intent are established as facts, the court must balance them and determine as a matter of law whether the scales compel a conclusion that inequitable conduct occurred. Id. If the court reaches that conclusion, it must hold the patent claims at issue unenforceable.

Du Pont argues that Atlas committed inequitable conduct by failing to tell the examiner that the examples were "prophetic" and, hence, in misleading the examiner into believing that the examples were actually performed. However, the district court found that the examples were written in the present tense to conform with the PTO requirements on prophetic examples. Moreover, the district court found that all but one of the examples were based on actual experiments and only slightly modified to reflect the inventor's notion of the most effective formulation. Consequently, the district court found, there was no intent on the part of Atlas to mislead the PTO. Du Pont has not shown those findings to be clearly erroneous. Even if intent could be inferred, and if the examples were prophetic but not disclosed to the examiner as such, Du Pont has not shown the nondisclosure to have been material, i.e., important to an examiner in allowing the patent to issue.

Du Pont asserts that Atlas' conduct cannot be distinguished from that in Grefco, Inc. v. Kewanee Industries, Inc., 499 F.Supp. 844, 208 USPQ 218 (D. Del. 1980), aff'd without publ. opinion, 671 F.2d 495 (3d Cir. 1981). We disagree. In Grefco, the patentee, to convince the examiner of the invention's superiority, presented "test results" based on tests that it knew never occurred, told the examiner the invention had been successfully tested when it had twice failed, and withheld information about those failures from the examiner. Intent and materiality were clearly established in Grefco, and the court in weighing the two factors held that there was inequitable conduct. That is not true here.

Du Pont argues that Atlas did not disclose its numerous "failures" and that it "padded" the disclosure with emulsifiers it knew would not work. The district court, however, found that Du Pont failed to prove that any of the emulsifiers were inoperative and the court found that the evidence on the "failed" experiments was not dispositive. Du Pont has not shown any clear error on the part of the district court in those findings.

Du Pont also alleges inequitable conduct in Atlas not disclosing to the examiner its Aquanite gel, the commercial version of the invention of its Gehrig patent. Though the district court found Aquanite to be "pertinent," it found no intent in the nondisclosure because Atlas had disclosed the Gehrig patent to the examiner. Du Pont has not shown any clear error in that finding. Cf. Vandenberg v. Dairy Equipment Co., 740 F.2d 1560, 1568-69, 224 USPQ 195 (Fed. Cir. 1984). (Vandenberg disclosed a PX-15 device as prior art but failed to describe it as its own prior invention; the disclosure was held to be inconsistent with the intent necessary for inequitable conduct).

VII. Infringement

Literal Infringement

Determining infringement requires claim construction as a preliminary step. See, e.g., Fromson v. Advance Offset Plate, Inc., 720 F.2d 1565, 1569, 219 USPQ 1137, 1140 (Fed. Cir. 1983). If properly construed claims read on the infringing product, there is literal infringement. Id. at 1571, 219 USPQ at 1142.

Du Pont's blasting agents are water-in-oil emulsions containing water, ammonium nitrate, fuel oil, occluded gas, and an emulsifying agent. Unlike the claimed invention, Du Pont uses as the emulsifying agent sodium oleate, which is formed in situ by adding sodium hydroxide and oleic acid to the other emulsion ingredients. Sodium oleate is normally an oil-in-water emulsifying agent but in the environment of the Du Pont product (i.e., a high salt concentration leading to

phase inversion), the sodium oleate acts as a water-in-oil emulsifying agent. The Du Pont product, and the in situ process of forming it, are the subject of U.S. Patent No. 4,287,100, issued to Owen and assigned to Du Pont.

The district court construed the '978 claim term "water-in-oil type emulsifying agent" as excluding compounds that normally function as oil-in-water emulsifying agents, e.g., sodium oleate. That claim construction prompted the district court to find no literal infringement. Atlas does not contest that finding and, for purposes of appeal, we accept it and the underlying claim construction.

Doctrine of Equivalents

A product that does not literally infringe can infringe under the doctrine of equivalents. Designed to protect a patentee from an infringer who appropriates the invention but avoids the literal language of the claims, the doctrine allows a finding of infringement when the accused product and claimed invention perform substantially the same function in substantially the same way to yield substantially the same result. Graver Tank & Mfg. Co. v. Linde Air Products Co., 339 U.S. 605, 608-09, 85 USPQ 328, 330 (1950); Perkin-Elmer Corp. v. Computervision Corp., 732 F.2d 888, 900, 221 USPQ 669, 679 (Fed. Cir. 1984), cert. denied, 53 U.S.L.W. 3226 (U.S. Oct. 2, 1984). The district court found that the Du Pont products and the claimed invention are equivalent, and Du Pont has not shown that finding to be clearly erroneous.[2]

The district court's opinion clearly delineates the Graver Tank tripartite test of substantially the same function, way, and result but then, adopting an analysis found in Ziegler v. Phillips Petroleum Co., 483 F.2d 858, 870, 177 USPQ 481, 487 (5th Cir.), cert. denied, 414 U.S. 1079, 180 USPQ 1 (1973), focuses on the "function, purpose, and quality" of the emulsifying agents of Du Pont and the claimed invention. That focus, argues Du Pont, was wrong because it ignored the Graver Tank tripartite test. We disagree.

Though Graver Tank articulates the tripartite test of "function, way, and result," it also states that the doctrine of equivalence should not be the prisoner of a rigid formula. Moreover, Graver, which as here compared a claimed mixture with an accused mixture in which one ingredient of the claimed mixture was changed, stated:

> "Consideration must be given to the purpose for which an ingredient is used in a patent, the qualities it has when combined with the other ingredients, and the function which it is intended to perform."

Id. at 611, 85 USPQ at 331.

Such consideration makes sense. Where, as here, the accused product avoids literal infringement by changing one ingredient of a claimed composition, it is appropriate for a court to consider in assessing equivalence whether the changed ingredient has the same purpose, quality, and function as the claimed ingredient. If it does, the accused and claimed products should meet the Graver Tank tripartite test of "function, way, and result."

That the district court focused on the function, quality, and purpose of the emulsifying agents does not mean it ignored the basic tripartite test which it expressly referred to in the opinion. We infer from that express reference, and from the opinion as a whole, that the district court did in fact find that the "function, way, and result" test was satisfied. See ACS Systems, Inc. v. Montefiore Hospital, 732 F.2d 1572, 1582, 221 USPQ 929, 936 (Fed. Cir. 1984) (this court will infer findings that were obviously necessary to the court's opinion).

Du Pont argues that, because its emulsion product was patented after the '978 patent issued, its product avoids infringement by equivalence. According to Du Pont, "*so long as direct infringement is lacking*, the grant of a patent to an accused infringer constitutes a prima facie determination of non-equivalence and, accordingly, of non-infringement" (Du Pont's emphasis). Atlas disagrees. So do we.

Du Pont concedes that, if Atlas patents A + B + C and Du Pont then patents the improvement A + B + C + D, Du Pont is liable to Atlas for any manufacture, use, or sale of A + B + C + D because the latter directly infringes claims to A + B + C. Du Pont urges, however, that it is not liable for manufacture, use, or sale of patented improvement A + B + C', even though A + B + C' is "equivalent" to A + B + C. We reject Du Pont's attempted distinction. Whether Du Pont makes A + B + C + D or A + B + C', Du Pont has used the gist of Atlas' invention to devise a patentable composition. There is no compelling reason to hold Du Pont liable

[2] One of Du Pont's products includes aluminum, which is not present in representative claim 1. It is, however, present in dependent claim 14. Moreover, the addition of an ingredient by Du Pont does not necessarily avoid infringement of claim 1. See, e.g., Radio Steel & Mfg. Co. v. MTD Products, Inc., 731 F.2d 840, 848, 221 USPQ 657, 663-64 (Fed. Cir. 1984), cert. denied, 53 U.S.L.W. 3225 (U.S. Oct. 2, 1984); Amstar Corp. v. Envirotech Corp., 730 F.2d 1476, 1482, 221 USPQ 649, 653 (Fed. Cir. 1984).

for infringement in one instance but not the other.[3]

[2] We agree with Bendix Corp. v. United States, 199 USPQ 203 (Ct. Cl. Trial Div. 1978), aff'd, 600 F.2d 1364, 204 USPQ 617 (Ct. Cl. 1979). There the trial judge said that where defendant has appropriated the material features of the patent in suit, infringement will be found "even when those features have been supplemented and modified to such an extent that the defendant may be entitled to a patent for the improvement." 199 USPQ at 221-22. Though Du Pont argues that cases from other courts support a contrary result, we are not bound by those cases and in any event find them unpersuasive.[4]

More persuasive is the reasoning of Herman v. Youngstown Car Mfg. Co., 191 F. 579, 584-85 (6th Cir. 1911). After finding equivalence, the court rejected appellant's contention that its receipt of a patent negates infringement:

> A patent is not the grant of a right to make or use or sell. It does not, directly or indirectly, imply any such right. It grants only the right to exclude others. The supposition that a right to make is created by the patent grant is obviously inconsistent with the established distinctions between generic and specific patents, and with the well-known fact that a very considerable portion of the patents granted are in a field covered by a former relatively generic or basic patent, are tributary to such earlier patent, and cannot be practiced unless by license thereunder.
>
> Another reason sometimes advanced for supposing that the structure of the second does not infringe the claim of the first patent is that the Patent Office has declared that a patentable difference exists. The premise is sound, but not the conclusion. In examining the second application, the Patent Office has no concern with the scope of the claim of the first, and does not and must not pay any attention thereto. It is concerned only with the early disclosure by the specification and drawings. Patentable difference does not of itself tend to negative infringement. It may just as well be based upon infringement, plus improvement; and improvement may lie in addition, simplification, or variance.

See also Sanitary Refrigerator Co. v. Winters, 280 U.S. 30, 42, 3 USPQ 40, 44 (1929) (where there is substantiality of function, way, and result, infringement cannot be avoided by any presumptive validity attaching to the issuance of a patent to the infringer); Sure Plus Mfg. Co. v. Kobrin, 719 F.2d 1114, 1117 (11th Cir. 1983) (no presumption of non-infringement arises from the issuance of a patent to the infringer); Freeman v. Altvater, 66 F.2d 506, 512, 18 USPQ 186, 192-93 (8th Cir.), cert. denied, 290 U.S. 696 (1933) (the court after finding equivalence stated that the issuance of a patent merely raises a presumption of validity, not a presumption of non-infringement).

Du Pont contends that one skilled in the art in 1966 would not have known that the '978 and Du Pont products were equivalent. It is not a requirement of equivalence, however, that those skilled in the art know of the equivalence when the patent application is filed or the patent issues. That question is determined as of the time infringement takes place. In Hughes Aircraft Co. v. United States, 717 F.2d 1351, 1365, 219 USPQ 473, 483 (Fed. Cir. 1983), this court held that devices changing the patented invention with advances developed subsequent to the patent could infringe under the doctrine of equivalents. See also American Hosp. Supply Corp. v. Travenol Labs., Inc., 745 F.2d 1, 9, 223 USPQ 577, 583 (Fed. Cir. 1984).

[3] Du Pont also argues that Atlas is "estopped" from asserting that the '978 claims cover the use of an oil-in-water emulsifier such as sodium oleate because Atlas was unable to use that type of emulsifier effectively. We reject Du Pont's argument on two grounds.

First, finding equivalence is not inconsistent with a patentee's unsuccessful attempt to make the accused product. The focus in assessing equivalence is on whether the accused product performs substantially the same as the claimed product in function, way and result — it is not on the patentee's ability to devise a product equivalent to the patented

---

[3] Of course, if A + B + C were patented because of unexpected results, those unexpected results might prompt a finding of no equivalence. That finding, however, would exist because, under the Graver Tank tripartite test, the "results" achieved by the claimed and accused products would be substantially different. The district court in this case did not find any such unexpected results. Though it found that Du Pont's products were more stable than those of the '978 patent, that is not necessarily inconsistent with equivalence. Equivalence does not require that the claimed invention and accused product have identical results; the results can be substantially the same and the accused product can be an improvement. Perkin-Elmer Corp. v. Computervision Corp., 732 F.2d at 901-02, 221 USPQ at 679-80; Decca Ltd. v. United States, 544 F.2d 1070, 1080-81, 191 USPQ 439, 448 (Ct. Cl. 1976).

[4] We are bound by opinions of our predecessor courts, the Court of Claims and CCPA. South Corp. v. United States, 690 F.2d 1368, 215 USPQ 657 (Fed. Cir. 1982).

product. Indeed, the patentee's incentive to devise an equivalent product is often less than a competitor's, which alone may account for the competitor's success and the patentee's failure in devising the equivalent product. See, e.g., Loesona Corp. v. Varta Batteries, Inc., 522 F.Supp. 1304, 1328, 213 USPQ 222, 241 (S.D.N.Y. 1981).

Second, the record submitted to this court makes no reference to any type of estoppel. That strongly suggests that estoppel was not raised before the district court. Borkoven v. Marsh, 727 F.2d 1558, 1566 (Fed. Cir. 1984). Because a party may generally not argue on appeal an issue not raised below, Weinar v. Rollform Inc., 744 F.2d 797, 804, 223 USPQ 369, 372 (Fed. Cir. 1984); Underwater Devices Inc. v. Morrison-Knudsen Co., 717 F.2d 1380, 1388, 219 USPQ 569, 575 (Fed. Cir. 1983), the estoppel argument is not properly before us.

Du Pont also argues that, because its product is formed in situ, it is different from the claimed product. It is the claimed product, however, not the process of forming it, that is involved. The district court found that the Du Pont emulsion, though it uses what is normally an oil-in-water emulsifier, "acts as a water-in-oil emulsifier," "caus[ing] a water-in-oil emulsion to form," and is otherwise substantially the same as the '978 emulsion. Those findings have not been shown to be clearly erroneous.

Du Pont further contends that the district court erred in considering the "heart of the invention" in its infringement analysis. We disagree. Although there is no legally recognized "essence" or "heart" of the invention in determining validity, W. L. Gore & Associates, Inc. v. Garlock, Inc., 721 F.2d 1540, 220 USPQ 303 (Fed. Cir. 1983), cert. denied, 53 U.S.L.W. 3226 (U.S. Oct. 2, 1984), it can be applicable in a determination of infringement under the doctrine of equivalents. Medtronic, Inc. v. Cardiac Pacemakers, Inc., 721 F.2d 1563, 1567, 220 USPQ 97, 101 (Fed. Cir. 1983). Moreover, the district court's "heart of the invention" analysis was supplemental to its finding that the Graver Tank tripartite test was satisfied.

Finally, Du Pont argues that the district court erred in not addressing in its opinion which of the individual claims are infringed. However, the district court specified the infringing claims in its judgment, and we review judgments, not statements in opinions. See, e.g., Lindemann Maschinenfabrik GMBH v. American Hoist & Derrick Co., 730 F.2d at 1463, 221 USPQ at 489. Reviewing the judgment, we conclude that the district court did not commit clear error in finding infringement of the claims on appeal.

VIII. Conclusion

Having considered all of Du Pont's arguments, the district court's decision that the '978 patent claims on appeal (1-5, 7, 12-14, and 16-17) are not invalid under 35 U.S.C. §§102, 103, and 112, that there was no inequitable conduct before the PTO, and that the claims on appeal were infringed, is *affirmed*.

Court of Appeals, Federal Circuit

State Industries, Inc. v. A.O. Smith Corporation

No. 84-590

Decided Jan. 3, 1985

PATENTS

1. Accounting — Increased or trebled damages or profits (§11.35)

Keeping track of competitor's products and designing new, better, or cheaper functional equivalents should not be discouraged by punitive damage awards except in cases where conduct is so obnoxious as clearly to call for them.

2. Notice and marking patented (§46)

Patent must exist and one must have knowledge of it for patent to be willfully infringed, and "patent pending" notice does not give one such knowledge and is not even guarantee that application has been filed, nor is filing guarantee that any patent will issue.

Particular patents — Water Heaters

4,263,879, Lindahl, Water Heater, holding of validity and infringement affirmed.

Appeal from District Court for the Middle District of Tennessee, Wiseman, J.; 221 USPQ 958.

Action by State Industries, Inc., against A.O. Smith Corporation, for patent infringement, in which defendant counterclaims for declaration of patent invalidity and noninfringement. From judgment for plaintiff, defendant appeals. Affirmed in part, and reversed in part.

Glen O. Starke, and Andrus, Sceales, Starke & Sawall, both of Milwaukee, Wis. (Gary

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants:	Gregory et. al.	Examiner:	Dr. K. C. Carlson
Serial No.:	08/087,132	Art Unit:	1812
Filed:	July 2, 1993	Docket:	NZI-012CN IG4-9.2(FWC)

For: NEW DIAGNOSTIC AND TREATMENT METHODS INVOLVING THE CYSTIC FIBROSIS TRANSMEMBRANE REGULATOR

Honorable Commissioner of
Patents and Trademarks
Washington, DC 20231

EXHIBIT BOOK — PART 2 of 2

S I R:

This Exhibit Book is submitted in connection with Applicants' Reply with Amendment of June 9, 1995 to the Offical Action that is pending herein

The Documents bound herein are:

(A) J. M. Rommens et al., "Identification of the Cystic Fibrosis Gene: Chromosome Walking and Jumping", Science, 245, 1989, pp. 1059-1065.

(B) J. Riordan et al., "Identification of the Cystic Fibrosis Gene: Cloning and Characterization of Complementary DNA", Science, 245, 1989, pp. 1066-1073.

(C) L. Tsui et al., "Cystic Fibrosis Gene", International Patent Application Number PCT/CA90/00267 published on March 7, 1991, bearing Publication Number WO 91/002796, and claiming the priority of United States patent applications 07/396,894, 07/399,945, and 07/401,609 filed, respectively, on August 22, 24, and 31, 1989.

(D) L. Tsui. et al. "Introns and Exons of the Cystic Fibrosis Gene and Mutations at Various Positions of the Gene", International Patent Application Number PCT/CA91/00009 published on July 25, 1991, bearing Publication Number WO 91/10734, and claiming the priority of Canadian national applications 2,007,699, 2,011,253, and 2,020,817 filed, respectively, on January 12, March 1, and July 10, 1990.

(E) F.S. Collins et al.,"Gene Therapy for Cystic Fibrosis", International Patent Application Number PCT/US91/06660 published on April 2, 1992, bearing Publication Number WO 92/05273, and claiming the priority of United States patent application 07/584,275 filed on September 18, 1990.

(F) United States Patent 5,240,846 to Collins et al. issued August 31, 1993 from Application No. 07/584,275 filed September 18, 1990

(G) L. Tsui et al., "Stable Propagation of Modified Full Length Cystic Fibrosis Transmembrane Conductance Regulator Protein cDNA in Heterologous Systems", International Patent Application Number PCT/CA91/00341 published on April 2, 1992, bearing Publication Number WO 92/05252, and claiming the priority of Great Britain national application 9020632.7 filed September 21, 1990.

(H) M. Drumm et al., "The Full Length CFTR cDNA is Toxic in Bacteria", Pediatric Pulmonology , Supplement 5 (Abstracts), October 1990, Abstract No. 8, page 189.

(I) M.L. Drumm et al., "Correction of the Cystic Fibrosis Defect in vitro by Retrovirus-Mediated Gene Transfer", Cell, 62, September 21, 1990, pp. 1227-1233.

(J) L. Tsui. "Probing the Basic Defect in Cystic Fibrosis", Current Opinion in Genetics and Development , 1, 1991, pp. 4-10.

(K) Ex parte Gray, Board of Patent Appeals and Interferences, 10 USPQ 2d 1929, 1989.

(L) Ex parte Forman, Board of Patent Appeals and Interferences, 230 USPQ 546, 1986.

(M) In re Wands, Court of Appeals, Federal Circuit, 8 USPQ 2d 1400, 1988.

(N) Ex parte Mark, Board of Patent Appeals and Interferences, 12 USPQ 2d 1904, 1989.

(O) Atlas Powder Co. v. E.I. DuPont Nemours, 750 F.2d 1569, 224 USPQ 409 (Fed. Cir. 1984).

Respectfully submitted,
GENZYME CORPORATION

Dated: 6/9/95

By: William G. Gosz
William G. Gosz, Esquire
Registration No. 27,787
Attorney for Applicants

One Mountain Road
Framingham, MA 01701- 9322
(508) 872-8400

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :		(11) International Publication Number:
C12P 21/06, C12N 5/00, 15/00 A61K 49/00, C12Q 1/68	A1	WO 92/05273
		(43) International Publication Date: 2 April 1992 (02.04.92)

(21) International Application Number: PCT/US91/06660

(22) International Filing Date: 16 September 1991 (16.09.91)

(30) Priority data:
584,275 18 September 1990 (18.09.90) US

(71) Applicant: THE REGENTS OF THE UNIVERSITY OF MICHIGAN [US/US]; 475 East Jefferson, Room 2354, Ann Arbor, MI 48109-1248 (US).

(72) Inventors: COLLINS, Francis S. ; 4340 Scio Church Road, Ann Arbor, MI 48103 (US). WILSON, James, M. ; 3686 River Pine Drive, Ann Arbor, MI 48103 (US).

(74) Agents: LEWAK, Anna, M. et al.; Harness, Dickey & Pierce, P.O. Box 828, Bloomfield Hills, MI 48303 (US).

(81) Designated States: AT (European patent), BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: GENE THERAPY FOR CYSTIC FIBROSIS

(57) Abstract

The present invention comprises gene therapy for treating cystic fibrosis (CF). Delivery and expression of a single copy of a normal CFTR gene leads to stable correction of the Cl channel regulation defect present in CF epithelial cells. The present invention includes recombinant viral and plasmid vectors, alternative CFTR gene delivery strategies, and transduced CF cells and cell lines carrying a recombinant gene for functional CFTR. CF epithelial complementation through transduction of the present invention also provides an assay for determining the validity of other putative CF mutations.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland			SN	Senegal
CI	Côte d'Ivoire	KR	Republic of Korea	SU+	Soviet Union
CM	Cameroon	LI	Liechtenstein	TD	Chad
CS	Czechoslovakia	LK	Sri Lanka	TG	Togo
DE*	Germany	LU	Luxembourg	US	United States of America
DK	Denmark	MC	Monaco		

\+ Any designation of "SU" has effect in the Russian Federation. It is not yet known whether any such designation has effect in other States of the former Soviet Union.

GENE THERAPY FOR CYSTIC FIBROSIS

SPONSORSHIP

Work on this invention was supported by the Cystic Fibrosis Foundation and by the United States government under grants DK42718 and DK39690 awarded by the National Institutes of Health. The government has certain rights in the invention.

RELATED APPLICATIONS

This an International Application, claiming benefit of U.S. application Serial No. 584,275, entitled "Gene Therapy for Cystic Fibrosis", filed September 18, 1990, which is a continuation-in-part of U.S. application Serial No. 401,609, entitled "Cystic Fibrosis Gene", filed on August 31, 1989, which is a continuation-in-part of U.S. application Serial No. 399, 945, entitled "Cystic Fibrosis Gene", filed on August 24, 1989, now abandoned, which is a continuation-in-part of U.S. application Serial No. 396,894, entitled "Cystic Fibrosis Gene", filed on August 22, 1989, now abandoned, all of which applications are specifically incorporated by reference herein.

FIELD OF THE INVENTION

The present invention relates generally to gene therapy for cystic fibrosis (CF) and, more particularly, to transfer of the gene for cystic fibrosis transmembrane conductance regulator (CFTR) to correct the defect in epithelial cell Cl channel regulation in cystic fibrosis patients.

BACKGROUND OF THE INVENTION

Cystic fibrosis (CF) is an autosomal recessive disorder characterized by abnormalities in water and electrolyte transport that lead to pancreatic and pulmonary insufficiency. Taussig, L.M., An overview. In Cystic Fibrosis, L.M. Taussig, ed. (New York: Thieme-Stralton), 1-9 (1984). It is one of the most common severe autosomal recessive disorders, having a 5% carrier frequency and affecting about 1 in 2500 live births in North America.

Functional expression of the CF defect reduces the chloride ion permeability of epithelial tissues. Quinton, P.M., *FASEB J.* 4:2709-2717 (1990). The ability of epithelial cells in the airways, sweat glands, pancreas and other tissues to secrete Cl in response to cAMP-mediated agonists is lost or severely reduced. Activation of apical membrane Cl channels by cAMP-dependent protein kinase (PKA) is impaired, but channels with normal conductance properties can be activated by other means, including agonists whose effects are mediated by increased cell Ca. Frizzell, R.A. et al., *Trends Neurosci.* 10:190-193 (1987); Welsh, M.J., *FASEB J.* 4:2718-2725 (1990). These findings suggest that the Cl channel *per se* is not defective in CF, but that the

defect might lie in a regulatory protein that transduces the effects of protein kinase activation. The presence of abnormalities in epithelial sodium transport in CF cells further supports the concept of a regulatory defect that can affect other cellular functions. Boucher, R.C. et al., *J. Clin. Invest.* 78:1245-1252 (1986).

Isolation of the gene for CF, as described in detail in the aforementioned related applications, has provided further insight into the molecular basis of the disease. See also Rommens, J.M. et al., *Science* 245:1059-1065 (1989); Riordan, J.R. et al., *Science* 245:1066-1073 (1989); Kerem, B.S. et al., *Science* 245:1073-1080 (1989). The gene responsible for CF has been localized to 250,000 bp of genomic DNA based on its location within the genome. This gene encodes a protein of 1480 amino acids called the cystic fibrosis transmembrane conductance regulator (CFTR). Riordan et al., *supra*.

The most compelling evidence thus far to support the role of CFTR in the etiology of CF has been provided by genetic analyses. Kerem et al., *supra*, (1989). Sequence analysis of the CFTR gene of CF chromosomes has revealed a variety of mutations, including nonsense and frameshift mutations. Cutting, G.R. et al., *Nature* 346:366-369 (1990); White, M.B. et al., *Nature* 344:655-667 (1990); Dean, M. et al., *Cell* 61:863-870 (1990); Kerem, B.S. et al., identification of mutations in regions corresponding to the 2 putative nucleotide (ATP) binding folds of the cystic fibrosis gene, *PNAS* (USA) (1990) (in press). However, extensive population studies have indicated that the most common CF mutation is a deletion of the three nucleotides that encode phenylalanine 508 (ΔF_{508}). This deletion is present on 70% of all CF chromosomes, but not on normal chromosomes. Kerem et al., *supra*, (1989); The Cystic Fibrosis Genetic Analysis Consortium (1990).

Results from both physiological and molecular cloning studies have raised the possibility that CFTR is a Cl channel. The defect in Cl channel activation by cAMP-dependent protein kinase (PKA) is present at the single-channel level in cell-free membrane patches and the protein structure predicted from CF gene cloning suggests that CFTR is an integral membrane protein with twelve membrane-spanning domains. Schoumacher, R.A. et al., *Nature* 330:152-754 (1987); Li, M. et al., *Nature* 331:358-360 (1988); Riordan et al., *supra*. The identification of CF-associated alterations in other cellular processes such as amiloride-sensitive Na transport and mucin sulfation also supports the view that CFTR may regulate several cellular processes. Boucher et al., *supra*; Boat, T.F. et al; *Arch. Biochem. Biophys.* 177:95-104 (1976).

Although the specific role that CFTR plays in Cl transport remains to be determined, the CFTR protein contains several interesting functional domains including

two nucleotide binding folds, a regulatory region that has many possible sites for phosphorylation, and two hydrophobic regions that probably interact with cell membranes. CFTR shows structural similarity with several members of the "ATP binding cassette" (ABC) superfamily of proteins, including the periplasmic binding proteins of prokaryotes, and the P-glycoprotein associated with multidrug resistance in higher eukaryotes. Riordan et al., *supra*; Hyde, S.C. et al., *Nature* 346:312-365 (1990).

Recent progress in our understanding of the genetic and functional basis of CF has provided a foundation for better defining its molecular pathology as well as developing novel therapies based on somatic gene transfer.

SUMMARY OF THE INVENTION

Gene therapy for cystic fibrosis (CF) comprises the delivery of a gene for functional cystic fibrosis transmembrane conductance regulator (CFTR) to affected epithelial cells. Delivery and expression of a single copy of the normal CFTR gene, as would be expected with a recessive trait such as CF, alleviates the Cl channel regulation defect present in CF cells. CF caused by a lack of functional CFTR or presence of CFTR function below physiologically-acceptable levels which arises from a defect in the CFTR gene can thus be treated in accordance with the principles of the present invention.

The "normal CFTR gene" of the present invention is simply any nucleic acid sequence which codes for functional CFTR. Thus variations in the actual sequence of the gene can be tolerated provided that functional CFTR can be expressed. For example, silent mutations can be introduced to stabilize cloning of the gene. A CFTR gene used in the practice of the present invention can be obtained through conventional methods such as DNA cloning, artificial construction or other means. The 4.6 kb cDNA utilized in the specific examples has all the sequences necessary to encode for a functional CFTR protein, as assayed by the analysis of cAMP-stimulated chloride current.

Gene transfer of the CFTR gene in accordance with the present invention can be accomplished through many means, including transfection using calcium phosphate coprecipitation, fusion of the target cell with liposomes, erythrocyte ghosts or spheroplasts carrying the CFTR gene, plasmid and viral vector-mediated transfer and DNA protein complex-mediated gene transfer.

Presently the delivery vehicle of choice is a recombinant retrovirus capable of infecting human epithelial cells. This is somewhat surprising given the relatively large

size of the CFTR gene. The recombinant retroviral vector of the invention generally comprises DNA of at least the portion of the retroviral genome necessary for infection, and the normal CFTR gene operatively linked thereto. Additionally, the portion of retroviral genome used in construction of the vector can be rendered replication-defective to remove any deleterious effects of viral replication on the target cells.

Although any CF-affected epithelial cells such as pancreatic and sweat gland cells can be targeted with the gene transfer methods and vectors of the present invention, because the most severe complications of CF are usually pulmonary, airway epithelial cells are the most desirable targets for gene therapy of the present invention. Moreover, given that airway epithelial cells have been found to be easily infected by recombinant retroviruses, gene transfer in accordance with the present invention to these cells is quite feasible.

CF diagnosis and screening of carriers can also be accomplished through transduction of CFTR defective cells and cell lines. For example, the complementation scheme of the present invention can be use to determine the validity of other putative CF mutations and is also useful as a tool to study the function of CFTR by site-specific mutagenesis or domain switching with other members of this large ABC gene family.

The present invention is thus directed toward gene therapy for cystic fibrosis through delivery and expression of a functional CFTR gene to the cells of a CF patient. Recombinant retroviral vectors as well as other CFTR gene transfer schemes can be used in the practice of the present invention. The present invention further includes both CF epithelial cells and cell lines which carry a normal CFTR gene transducted or transferred therein in accordance with the principles of the invention. CFTR screening and complementation assays for other putative CF mutations are also contemplated within the scope of the present invention.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1A depicts the proviral component of the recombinant retroviral vector PLJ-CFTR of the present invention.

Figure 1B shows the results of gel blot hybridization of Kpn I-restricted DNA using a Neo specific probe (top panel) and Hind III digested DNA using an exon 13 CFTR specific probe (bottom panel).

Figure 2 is an RNA blot analysis of the expression of the retroviral transduced CFTR gene in CFPAC clones using the CFTR exon 13 probe.

Figure 3A is a graph illustrating the effect of forskolin on ^{125}I efflux in PLJ and PLJ-CFTR clones over time.

Figure 3B is a graph illustrating the effect of forskolin on ^{125}I efflux in PLJ and PLJ-CFTR clones as compared to basal ^{125}I efflux.

Figure 4A is a whole-cell voltage clamp record of stimulation of inward currents by forskolin in a PLJ-CFTR clone 1.

Figure 4B illustrates the failure of cAMP or forskolin in stimulating membrane currents in a PLJ clone 6 cell.

Figure 4C is a graph depicting the instantaneous current-voltage relations of forskolin-induced currents in NaCl, low Cl and Na-free baths.

Figure 5 is the nucleotide sequence of cDNA encoding CFTR together with the deduced amino acid sequence.

Figure 6 depicts a stabilization scheme for a CFTR construct.

Figure 7 is a restriction map of a plasmid-based vector used in the practice of the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS OF THE INVENTION

The absence of functional CFTR or CFTR function which is not at physiologically-acceptable levels and which arises from a defect in the CFTR gene is treated through gene transfer of a normal CFTR gene into CFTR defective cells. By "physiologically-acceptable level of CFTR function" is meant a level of CFTR function at which a cell population or patient exhibits the normal physiological effects presence of the normal amounts of CFTR. Examples of insufficiencies in CFTR function include but are not limited to abnormal Cl channel regulation in epithelial cells, such as that exhibited in cystic fibrosis.

A recombinant viral vector of the present invention comprises DNA of at least a portion of retroviral genome which portion is capable of infecting the target cells and a normal CFTR gene operatively linked to thereto. By "infection" is generally meant the process by which a virus transfers genetic material to its host or target cell. Preferably the retrovirus used in the construction of a vector of the invention is also rendered replication-defective to remove the effects of viral replication on the target cells. In such cases, the replication-defective viral genome can be packaged by a helper virus in accordance with conventional techniques. Generally any retrovirus meeting the above criteria of infectiousness and capabilities of CFTR gene transfer can be employed in the practice of the present invention may also be desirable. Suitable retroviruses for the practice of this invention include, for example, PLJ, pZip, pWe and

pEM well known to those skilled in the art. Suitable packaging virus lines for replication-defective retroviruses include, for example, ΨCrip, ΨCre and Ψ2 and ΨAm.

It will be appreciated that when viral vector schemes are employed for CFTR transfer, the use of attenuated or avirulent viruses may also be desirable. Where applicable in the practice of the invention, amplification of the CFTR gene can also be utilized to enhance the levels of normal CFTR expression.

The genetic material to be recombined with the retroviral vector or transferred through other methods of the invention is preferably provided through conventional cloning methods, i.e. cDNA, through overlapping oligonucleotide sequences or any other suitable method yielding the desired sequence. When used in diagnostic or screening assays, the genetic material is usually provided by cloning of patient DNA or, alternatively, through the use of patient genomic DNA. As stated previously, by normal CFTR gene, is meant any nucleic acid sequence which codes for functional CFTR.

The cells targeted for transduction or gene transfer in accordance with the present invention include any cells to which the delivery of the CFTR gene is desired. Generally speaking, the cells are those with the CFTR gene defect, such as CF cells. In the case of CF, the cells targeted are preferably epithelial cells, including pancreatic, sweat gland, liver, intestinal, kidney and even more preferably epithelial airway cells, such as lung cells.

Cells or cell populations can be treated in accordance with the present invention *in vivo* or in *in vitro*. For example, in *in vivo* treatments, CFTR vectors of the present invention can be administered to the patient, preferably in a biologically compatible solution or pharmaceutically acceptable delivery vehicle, by ingestion, injection, inhalation or any number of other methods. The dosages administered will vary from patient to patient and will be determined by the level of enhancement of CFTR function balanced against any risk or deleterious side effects. Monitoring levels of transduction, CFTR expression and/or the presence or levels of normal CFTR will assist in selecting and adjusting the dosages administered. *in vitro* transduction is also contemplated within the present invention. Cell populations with defective CFTR genes can be removed from the patient or otherwise provided, transduced with a normal CFTR gene in accordance with the principles of the invention, then (re)introduced into the patient.

CFTR-defective cell lines, such as transformed CF lines, can also be transduced in accordance with the present invention. Such cell lines are useful, for example, in

complementation assays for evaluating CF mutations to diagnose CF and screen for carriers. For example, patient CFTR cDNA can be transferred into CF cells and the cells screened for complementation, i.e. CFTR function, to confirm or rule out of CF or CFTR gene defects.

In the first set of Specific Examples which follow, retrovirus-mediated gene transfer was used to complement the cystic fibrosis (CF) defect in Cl regulation in epithelial cells of a CF patient. Amphotropic retroviruses were used to transduce a functional cystic fibrosis transmembrane conductance regulator (CFTR) cDNA into CFPAC-1, a pancreatic adenocarcinoma cell line derived from a patient with CF. This cell line stably expresses abnormalities in electrolyte transport that are characteristic of the CF defect, i.e. they lack cAMP-stimulated Cl transport. CFPAC-1 cells were exposed to control virus (PLJ) and CFTR-expressing virus (PLJ-CFTR); viral transduced clones were isolated and subjected to molecular and physiologic analysis. Agarose gel blot analysis revealed unrearranged proviral sequences in 10 of 10 PLJ clones and 9 of 10 PLJ-CFTR clones. RNA analysis detected a viral-derived CFTR transcript in all of the PLJ-CFTR clones. RNA analysis detected a viral-derived CFTR transcript in all of the PLJ-CFTR clones that contained unrearranged proviral sequences.

Anion (^{125}I) efflux was used to examine PLJ and PLJ-CFTR clones for cAMP and Ca stimulated anion transport. Agents that increase intracellular cAMP stimulated ^{125}I efflux in PLJ-CFTR clones but not PLJ clones. Whole-cell patch-clamp performed on two responding clones showed that the anion efflux responses were due to cAMP stimulation of chloride conductance. Calcium ionophore increased ^{125}I efflux and chloride currents in all PLJ and PLJ-CFTR clones. These findings indicate that expression of the normal CFTR gene confers cAMP-dependent Cl channel regulation on CF epithelial cells.

The second set of Specific Examples describes gene transfer to airway epithelial cells as well as pancreatic cells, administration of CFTR gene therapy and alternative gene transfer delivery systems, including lipofection and DNA-protein complex- mediated gene transfer.

SPECIFIC EXAMPLES - I

Recombinant Retroviruses

Early attempts to reconstitute a full length CFTR cDNA from overlapping clones were unsuccessful. The exact cause of these difficulties remains to be defined, but our data indicated that prokaryotic transcription from internal CFTR cDNA sequences may result in the expression of a protein that is toxic to bacteria. The introduction of three

silent mutations (T to C at 930, A to G at 933, and T to C at 936) into a restriction fragment of CFTR that spans exon 6b completely ablated this toxic effect, potentially by interfering with the cryptic prokaryotic promoter, and enabled the reconstruction of 4.6 kb of contiguous CFTR cDNA sequence. The nucleotide sequence of this reconstructed cDNA was re-determined and found identical to that published previously with the exception of the three silent mutations noted above. See Riordan et al., *supra*.

Sequence Listing set forth before the Claims illustrates the nucleotide sequence of cDNA encoding CF transmembrane conductance regulator along with the deduced amino acid sequence. The DNA sequencing was performed by the dideoxy chain termination method with ^{35}S-labeled nucleotides or by the Dupont Genesis 2000 automatic DNA sequencer. Numbers on the right of the columns in Figure 5 indicate base positions. The first base position corresponds to the first nucleotide in the 5′ extension clone PA3-5, which is one nucleotide longer than TB2-7. The 3′ end and the noncoding sequence are also shown in the Figure (nucleotides 4561 to 6129 plus the poly(A)$^+$ tail). The arrows shown indicate the position of transcription initiation site by primer extension analysis. As shown abbreviated in the Figure, nucleotide 6129 is followed by a poly(A) tract. The positions of exon junctions are between the following base positions: 185-186; 296-297; 405-406; 621-622; 711-712; 1001-1002; 1248-1249; 1341-1342; 1523-1524; 1716-1717; 1811-1812; 1898-1899; 2621-2622; 2789-2790; 3040-3041; 3120-3121; 3499-3500; 3599-3600; 3849-3850; 4005-4006; 4095-4096; 4268-4269; and 4374-4375.

Potential membrane-spanning segments were ascertained with the use of the algorithm of Eisenbert, et al., *J. Mol. Biol.* 179:125 (1984) and are underlined under the nucleic acid designations in the Figure. Amino acids comprising putative ATP-binding folds are underlined under the animo acid designations in the Figure. Possible sites of phosphorylation by protein kinases A or C are indicated by o and *, respectively, and + designates glycosylation sites. The open triangle indicates the position at which 3 bp are deleted in CF. Abbreviations for the amino acid residues used in Figure 5 are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gin; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Try.

The modified CFTR cDNA was cloned into the retroviral vector PLJ previously described by Korman, A.J. et al., *PNAS (USA)* 84:2150-2154 (1987). The proviral component of this recombinant vector, called PLJ-CFTR, is depicted in Figure 1A. Important structural components of the vector illustrated in the Figure include the long-terminal repeat sequences (LTR), CFTR cDNA, sequences from the origin of SV40, the

gene that confers resistance to G418 (Neo), and the origin of replication for pBR322 (pBR). Transcription from the 5'LTR produces the 8.5 kb genomic transcript that is responsible for viral passage and CFTR expression. Transcription from the SV40 sequences leads to the formation of the second transcriptional unit that expresses a Neo-selectable marker. Sites of transcriptional initiation are indicated in Figure 1A with arrows at the 5' LTR and internal SV40 sequences. Recognition sites for restriction endonucleases Kpn I and Hind III are also indicated. Probes specific for the CFTR gene (exon 13) and the Neo gene that were used in DNA and RNA blot analysis are indicated below the vector.

Transfection of PLJ and PLJ-CFTR vectors into the virus packaging cell line ΨCrip led to the transient production of replication-defective virus. Limiting dilutions of virus stocks were used to infect CFPAC-1 cells which were subsequently cultured in the presence of G418 in order to select for transduced clones. Transiently produced PLJ-CFTR virus stocks had a lower titer (50-100 fold) than those produced with PLJ vector. Ten individual clones of cells were isolated from infections performed with each type of virus (named PLJ clones 1 through 10 and PLJ-CFTR clones 1 through 10) and subjected to molecular and physiologic analysis.

Transduced Clones Express Retroviral CFTR Sequences

Retrovirally transduced clones of CFPAC-1 cells were analyzed for proviral sequences as described for other cells types by Wilson, J.M. et al., *PNAS (USA)* 85: 4421-4425 (1988) and Wilson, J.M. et al., *Science* 248:1413-1416 (1990). CFPAC-1 cells were infected with PLJ or PLJ-CFTR virus and selected in the presence of G418 in order to isolate individual clones. High molecular weight DNA was isolated from each clone and analyzed by the method of gel blot hybridization as shown in Figure 1B. In the top panel of Figure 1B, DNA was digested with Kpn I and the filter was hybridized to a Neo specific probe, whereas in the bottom panel, DNA was digested with Hind III and the filter was hybridized to the exon 13 CFTR specific probe. The 4.3 kb band in all lanes arises from the endogenous CFTR gene. Samples include: CFPAC-1 DNA (10 μg); lane "1 copy" - CFPAC-1 DNA (10 μg) and lanes "PLJ-CFTR 1 through 10" - DNA (10 μg) from PLJ-CFTR clones 1 through 10. Along the right border of the figure, molecular size standards in kilobases are indicated.

Digestion of high molecular weight DNA with restriction enzyme Kpn I, which has unique sites in the vector LTRs, released all integrated forms of the PLJ-CFTR provirus as a common 8.5 kb fragment. As indicated in the top panel of Figure 1B, gel blot hybridization of Kpn I-restricted DNA revealed unrearranged proviral sequences

with the expected abundance of one copy per cell in 10/10 PLJ clones and 9/10 PLJ-CFTR clones. Hybridization of the filter with a Neo-specific probe detected a markedly rearranged provirus in PLJ-CFTR clone 2; this virus apparently deleted a major part of the CFTR cDNA (data not shown).

The results of gel blot hybridization analysis to study the complexity and uniqueness of each putative PLJ-CFTR clone is shown in the bottom panel of Figure 1B. High molecular weight DNA was isolated and digested with Hind III, a restriction enzyme with two internal sites in PLJ-CFTR, and analyzed with the exon 13 CFTR specific probe. As illustrated in the bottom panel of Figure 1B, this analysis demonstrated a single unique integration site in 9/10 PLJ-CFTR clones. The CFTR specific probe failed to detect the provirus in DNA from PLJ-CFTR clone 2 because of the apparent deletion described above.

Expression of the retroviral transduced CFTR gene was studied by RNA blot analysis of CFPAC-1 clones using the CFTR exon 13 probe and is shown in Figure 2. Clones of retrovirus transduced CFPAC-1 cells were isolated and analyzed for the presence of CFTR transcripts. Total cellular RNA was harvested from individual clones and subjected to RNA blot analysis using the exon 13 CFTR probe to hybridize with the filter as shown in the top panel of Figure 2. In the bottom panel of Figure 2, the filter was stripped and rehybridized with a probe derived from human γ-actin cDNA in order to control for variation in sample loading. RNA samples (10 μg) were derived from the following cells: lanes "T84" - duplicate samples from the colonic tumor cell line T84; lane "CFPAC-1" - nontransduced CFPAC-1 cells; lane "PLJ 6" - CFPAC-1 clone #6 from the PLJ infection; and lanes "PLJ-CFTR 1 to 10" - CFPAC-1 clones #1 through #10 from the PLJ-CFTR infection. Along the left border of the figure, molecular size standards in kilobases are indicated.

As illustrated in Figure 2, total cellular RNA from the previously described human colon tumor cell line, T84, demonstrated high levels of the endogenous CFTR transcript. No CFTR transcript was detected by Northern analysis in mock infected CFPAC-1 cells or PLJ clones 1 though 10 CFTR RNA can be detected in CFPAC-1 by RNA-PCR. A viral directed CFTR transcript of the expected size (i.e., 8.5 kb) was detected in 9/10 PLJ-CFTR clones; the CFTR probe failed to detect a transcript in RNA from the clone that contains the deleted provirus (PLJ-CFTR clone 2).

Transduced Clones Show Forskolin Stimulation of Anion Transport

Isotopic anion (^{125}I) effluxes were measured to screen the PLJ and PLJ-CFTR clones for cAMP- and Ca-stimulated anion transport. The efflux assay, described by

Venglarik, C.J. et al., *Am. J. Physiol.* 259:C358-C364, (1990) provides a qualitative estimate of agonist-stimulated Cl conductance pathways in Cl-secreting epithelia, as judged from the inhibitory effects of Cl channel blockers and depolarizing membrane potentials on ^{125}I efflux. Figure 3A shows the time-course of the ^{125}I efflux rate constant (r) in two clones, PLJ 6 and PLJ-CFTR 6, with and without the addition of forskolin, an agent which stimulates adenylate cyclase. 10 μM of forskolin was added at the time indicated in Figure 3A. Following a basal efflux period in the absence of agonist (not shown), forskolin increased ^{125}I efflux rate from PLJ-CFTR clone 6 from 0.32 to 0.70 min^{-1}; PLJ 6 did not respond. r values obtained before forskolin addition and during the peak of the forskolin response provided an estimate of the relative stimulation of ^{125}I efflux (i.e. r_{forsk}/r_{basal}). In the responding PLJ-CFTR clones, the peak forskolin effect on anion efflux was observed during the first three flux periods following forskolin addition (15-45 sec). The mean +/-SEM was n=9 for all clones except PLJ5 where n=7.

Data derived from twenty clones is illustrated in Figure 3B. The r values were taken before and after the addition of forskolin. For PLJ-CFTR clone 2 the same scaling applies below 1.0. The values are mean +/-SEM; n=9 for all clones except PLJ5 where n=7. As illustrated in Figure 3B, seven of ten PLJ-CFTR clones showed significant increases in ^{125}I efflux in response to forskolin, whereas none (0/10) of the control PLJ clones responded to forskolin. The parent cell line, CFPAC-1, also shows no response to forskolin or cAMP analogues as described by Schoumacher, R.A. et al., *PNAS (USA)* 87:4012-4016 (1990). PLJ-CFTR clone 2 showed a major deletion in its CFTR cDNA by gel blot hybridization as shown in Figure 1B, accounting for the failure of forskolin to stimulate ^{125}I efflux. In the seven responding PLJ-CFTR clones, the relative stimulation of anion efflux by forskolin ranged from 1.8 to 2.8-fold. This compares well with the 3.5-fold stimulation of efflux reported recently for the colonic tumor cell line T84 by Venglarik, *supra*. Our results indicate that expression of CFTR cDNA endows CFPAC-1 cells with cAMP-responsive anion efflux.

The correlation between forskolin responsiveness of the PLJ-CFTR clones and their CFTR mRNA levels was not striking as illustrated by a comparison of Figures 2 and 3B. Three of the best responders in efflux assay showed high mRNA levels (i.e., PLJ-CFTR clones 1, 6 and 10). In other instances, however, the correlation was not as good. For example, clones 7 and 8 showed approximately a 2-fold response to forskolin but had relatively low mRNA levels, and clones 3 and 9 showed a low forskolin response, despite the presence of readily detectable CFTR mRNA.

Addition of the Ca ionophore, ionomycin, increased ^{125}I efflux in all control and CFTR clones. Values of ^{r}iono/^{r}basal averaged 14+/_2 in PLJ and 14+/_1 in PLJ-CFTR (n=20) in each group; no significant differences were detected between individual clones. The extent of response of PLJ clones to ionomycin is similar to that observed previously in wild-type CFPAC-1 cells by Schoumacher et al., *supra* (1990), and is about three times the response of T84 cells observed by Venglarik et al., *supra*. The ability of Ca ionophores and Ca-mediated agonists to stimulate Cl secretion has been reported for airway and sweat gland cells derived from both normal individuals and CF patients. See Sato, K. et al., *J. Clin. Invest.* 73:1763-1771 (1984); Frizzell et al., *supra* (1986); Willumsen, N.J. et al., *Am. J. Physiol.* 256:C226-C233 (1989). The presence of this response in CF cells indicates that CFTR is not required for Ca-mediated Cl transport stimulation. The lack of significant differences in the extent of Ca stimulation in PLJ and PLJ-CFTR clones suggest that CFTR does not modulate the activity of Ca-mediated regulatory pathways that govern Cl secretion.

Clones Transduced with the CFTR Retrovirus Show cAMP-Induced Cl Currents

Whole-cell path-clamp recordings were used to determine whether the cAMP-induced increase in anion efflux in PLJ-CFTR clones of Figure 3 was due to stimulation of Cl conductance pathways as described by Cliff, W.H. et al., *PNAS (USA)* 87:4956-4960 (1990). A typical response of PLJ-CFTR clone 1 is shown in Figure 4A which illustrates stimulation of inward currents by 5μM forskolin. Membrane voltage was held at -10 mV and pulsed to 0 and -84 mV. The gap in the record represents time (6 min) during which bath solution substitutions were performed to determine ion selectivity of the forskolin-induced current. The pulse protocols for determining the I-V relations were run at the times indicated in the Figure. Chloride currents were measured as the inward current produced by voltage pulses to -84 mV. Similar increases in inward current were observed in 11 of 13 cells from PLJ-CFTR clones 1, 6, and 10 in which addition of forskolin (5 μM) or cAMP (200 to 800 μM) increased inward currents from 220 +/-68 pA to 1690 +/-495 pA in responding cells. The magnitude of this response compares favorably with that observed in T84 cells by Cliff et al., *supra*. As shown in Figure 4B, cAMP (400 μM) of forskolin (5 μM) failed to stimulate membrane currents in cells from the control clone, PLJ 6 (n=6). The membrane voltage was held at -20 mV and pulsed to 0 mV and -84 mV. Similar results were obtained in 5 PLJ clone 6 cells. As observed from the ^{125}I efflux determinations, ionomycin (2 μM) increased inward currents in both PLJ (n=4) and PLJ-CFTR (n=3) clones.

As shown in Figure 4C, instantaneous current-voltage (I-V) relations of the forskolin-stimulated current in a NaCl bath, a low Cl bath, and a Na-free bath were obtained from PLJ-CFTR clone 1. Forskolin-induced currents were obtained by digital substraction of currents before and after stimulation. The values shown in Figure 4C were recorded 6 msec after the initiation of voltage pulses. These data were obtained from the PLJ-CFTR clone 6 cell record shown in Figure 4A during the 6 min. recording gap.

As illustrated in Figure 4C, the I-V relation of the stimulated current appeared to be linear, as observed in T84 cells by Cliff et al., *supra*. Currents were determined using equal bath and pipette Cl concentrations reversed near the Cl equilibrium potential of 0 mV. Reducing bath Cl to 6 mM (glutamate replacement) decreased the outward currents and shifted the reversal potential for current flow to +66 mV, a value close to the Cl equilibrium potential (+80 mV) for this outwardly-directed Cl gradient. Replacement of bath Na by N-methyl-D-glucamine (NMDG) did not significantly alter the I-V relation. These findings indicate that the forskolin-stimulated current is Cl-selective, and that the stimulation of anion efflux in PLJ-CFTR clones is due to activation of Cl conductance pathways.

EXPERIMENTAL PROCEDURES

The following experimental procedures were employed in the Specific Examples set forth above:

CFPAC-1 cells were maintained in culture as described previously by Schoumacher et al., *supra* (1990); cells used for retroviral infection were at passage 72. Infection populations of CFPAC-1 cells were selected in medium containing G418 (1 mg/ml) in order to isolate individual clones. Transduced CFPAC-1 cells were removed from selection soon after they were expanded as clones. This was not associated with an apparent loss of proviral sequences or proviral expression. The amphotropic packaging cell line ΨCrip, was maintained in Dulbecco's modified Eagle's medium supplemented with 10% calf serum and penicillin/streptomycin as described by Danos, O. et al., *PNAS (USA)* 85:5460-6464 (1988).

Construction of CFTR cDNA

The cDNA was constructed by joining the overlapping clones 10-1, T16-1 and T16-4.5 as described by Riordan et al., *supra*. 10-1 and T16-1 were ligated at the unique Nru I *site in exon 4 and the resultant construct, spanning exons 1 thorough 13, joined to T16-4.5. This was done by inserting a Sac I-Eco RI* partial digestion product of T16-4.5, extending from exon 13 to exon 24, into the respective sites of the 5' 13-

exon construct. These manipulations generated a 4.5 kb clone containing the entire coding sequence as previously described by Riordan et al., *supra*. It was observed that most clones generated from these construction attempts were grossly rearranged. Upon sequencing of an apparently intact construct, a 57 bp deletion was identified in exon 6b occurring between the two copies of a 13 bp direct repeat. On inspection, this interval was noted to contain a consensus prokaryotic promoter sequence. In an attempt to disrupt the repeat, three single nucleotide alterations were made by *in vitro* mutagenesis. The introduced changes which do not alter the CFTR translation product and result in a stable construct, include substitution of C for T at position 930, G for A at 933, and C for T at position 936. The modified reconstructed CFTR plasmid is called CFTR 4.6.

The above described changes were accomplished by synthesis of an oligonucleotide which matched the normal sequence except for the presence of G at 933 and C at 936. The antisense strand of this segment of the CFTR cDNA was cloned into single-stranded M13 phage, and mutagenized with the oligonucleotide using standard techniques as described by Smith, M., (1989) *Annu. Re. Genet.* 19:423 (1985); Sanbrook, J., et al.: Molecular cloning. A Laboratory Manual, 2nd ed., Cold Spring Harbor Press, 15.51-15.80 (1989). The resulting clone, shown in Figure 6, was sequenced and found to have an additional unexpected base change at position 930, which is also in a silent nucleotide position not altering the encoded protein.

It will also be appreciated that other methods to stabilize the full-length CFTR cDNA can be used in the practice of the invention. Any alteration in the fortuitous E. coli promoter in exon 6b, as shown in Figure 6, which renders it non-functional while preserving the correct amino acid coding sequence for CFTR will accomplish this same goal. For example, mutagenesis of the CATACT sequence underlined in Figure 6 can be accomplished in several ways which will not alter the amino acid sequence (e.g. CGTATT), but will inactivate the sequence as a prokaryotic promoter, rendering it stable in the usual cloning vectors.

Retroviral Vectors and Recombinant Retroviruses

Digestion of the modified CFTR plasmid with Sac I released the modified CFTR cDNA on a 4.6 kb restriction fragment. The Sac I sites were converted to Bcl I sites with oligonucleotides and the linkered fragment was cloned into the Bam I site of the retroviral vector PLJ previously described by Korman et al., *supra*. This recombinant vector, called PLJ-CFTR, is presented in Figure 1A. Retroviral vector PLJ and PLJ-CFTR were transfected into the amphotropic packaging cell line ΨCrip as described.

Tissue culture media was removed from plates containing the transfected packaging cells 24 hours later in order to harvest the transiently produced amphotropic virus.

CFPAC-1 cells, passaged 1:5 onto 10 cm^2 plates, were exposed to viral supernatants supplemented with polybrene (4 μg/ml) for 12 to 16 hours. When the cells reached confluence, they were passaged 1:10 into medium containing G418 (1 mg/ml). Clones of cells were isolated, expanded, and cryopreserved.

DNA and RNA Analysis of CFPAC-1 Clones

High molecular weight DNA was isolated from CFPAC-1 cells as described and analyzed by gel blot hybridization as described by Wilson et al., *supra* (1988). Total cellular RNA was purified and subjected to RNA blot analysis of Wilson et al, *supra* (1988). Filters were hybridized with a variety of DNA probes that were labeled to a high specific activity using the random priming method of Feinberg, A.P. et al., *Anal. Biochem.* 132:6-13 (1983). These probes include: 1) Exon 13 of CFTR isolated following PCR amplification of cloned cDNA using oligonucleotides that flank the borders of this exon, (NT 1900 to 2611); 2) Neo-specific sequences on a 960 base pair Hind *III* to Nco I fragment of pSV2Neo, and 3) human γ actin cDNA.

Anion Efflux Measurements

Radioisotopic anion efflux was determined as described by Venglarik et al., *supra*. Briefly, cell monolayers were preloaded with ^{125}I for 30 min; after two washes, efflux was monitored at 15 sec intervals using a sample-replace procedure. At the end of the experiment, tracer remaining in the cell monolayer was extracted with 0.1 N HPO_3. The efflux rate constant (r) for each sampling interval was calculated as follows: $r = [\ln (R_1) - \ln (R_2)]/(t_1 - t_2)$, where R_1 and R_2 are the percent of loaded ^{125}I remaining in the monolayer at times (t) 1 and 2. Forskolin or ionomycin were added after the fifth 15-sec sampling interval. The degree of agonist stimulation is expressed as $r_{agonist}/r_{basal}$ where $r_{agonist}$ is the maximal value observed in the presence of agonist and r_{basal} is taken from flux interval immediately prior to agonist addition.

Most of the extracellular ^{125}I washout occurs during the initial 60 sec of sampling as set forth by Venglarik et al., *supra*; this period was ignored in the rate constant calculations. However, a small residual efflux from the extracellular space after 60 sec leads to a slight underestimate of the agonist response because the extracellular compartment washes out faster than the cellular compartment. Therefore, when there is no efflux response to forskolin, r determined immediately after forskolin addition is slightly less than that measured before forskolin is added. This accounts

for the finding that $\gamma_{forsk}/\gamma_{basal}$ is between 0.9 and 1.0 in the PLJ clones shown in Figure 3B.

Whole-Cell Current Recordings

Macroscopic currents were recorded during whole-cell patch-clamp by methods previously described by Cliff et al., *supra*. Recordings were made at 37°C with the following solutions (mM); bath: 115 NaCl, 40 N-methyl-D-glucamine (NMDG)-glutamate, 5 K-glutamate, 2 $MgCl_2$, 1 $CaCl_2$, 10 HEPES (pH 7.2); pipette: 115 KCl, 35 NMDG-glutamate, 0.25 EGTA, 0.09 $CaCl_2$ (100 nM free Ca), 2 $MgCl_2$, 2 Na_2ATP, 0.20 Na_2GTP, 10 HEPES (pH 7.2). Membrane potentials were clamped alternately for 500 msec duration at three voltages, two of which were chosen to equal the equilibrium potentials for Cl (0 mV) and K (_84 mV). This permits the Cl and K currents to be monitored during agonist responses as described by Cliff et al., *supra*. Pulsing was interrupted to determine current-voltage relations by stepping the clamp voltage between +/_100 mV at 20 mV increments as shown in Figure 4C.

SPECIFIC EXAMPLES - II

Retrovirus-Mediated Transduction of Pancreatic and Pulmonary Epithelial Cells

Retrovirus-mediated gene transduction into various epithelial cells was optimized using a replication defective retrovirus that expresses the β-galactosidase gene from E. coli. This was used because expression of viral directed β-galactosidase can be detected *in situ* using cytochemical reaction that stains the transduced cell blue. The amphotropic virus producer cell line made from the β-galactosidase expressing BAG vector, which has been described previously, was used as a source of virus. This virus producing cell line is called BAG5. The supernatant over a confluent plate of BAG5 cells was harvested, filtered, and used to infect various epithelial cells as described below.

Pancreatic Epithelial Cell Line

CFPAC-1 is a cell line derived from an adenocarcinoma of a patient with CF which expresses the cellular defect characteristic of CF (i.e. chloride channels are not activated in the presence of cAMP agonists). CFPAC-1 cells were split at various dilutions (1:2, 1:5, 1:10, and 1:20) and 24 hours later exposed to fresh virus supernatants that had been supplemented with polybrene (4 μg/ml). Twelve hours later the virus was replaced with fresh medium. When confluent, the cells were analyzed for the expression of viral directed β-galactosidase as described. Optimal infection efficiency was obtained with CFPAC-1 cells that were split 1 to 5 the day before infection. Under optimal conditions, a single exposure to virus led to stable

transduction of the β-galactosidase gene into 30-40% of the cells. Expression of β-galactosidase has been stable in cultured cells for over 2 months. Attempts to reinfect CFPAC-1 cells on subsequent days led to little augmentation of infection efficiency.

Airway Epithelial Cells

As discussed previously, airway epithelial cells are the most desirable targets for gene transfer because the pulmonary complications of CF are usually its most morbid and life-limiting. Taussig, *supra* (1984). Since airway epithelial cells are easily infected with recombinant retroviruses, the gene transfer approaches described in the preceding and following examples will also be useful for gene therapies directed to airway epithelial cells such as those of the lung.

An epithelial cell line derived from an airway of a patient with CF was used as a potential target for retrovirus-mediated gene transfer. These cells had been described previously and have been called T43 cells. Freshly harvested BAG5 virus was supplemented with polybrene and exposed to T43 cells that had been split 1:5, 24 hours previously. Cells were exposed to virus for 12-18 hours and allowed to grow to confluence before being analyzed for viral directed β-galactosidase expression using the previously described cytochemical assay. Under optimal conditions, greater than 25% of CFPAC cells were stably transduced with the β-galactosidase gene after a single exposure to virus.

Direct Delivery of CFTR Expressing Vectors to the Airway Epithelial Cells

One approach to the use of recombinant retroviruses and the treatment of CF is to introduce a functional CFTR gene into epithelial cells *in vivo* by directly delivering retroviruses into the airway. Several approaches can be taken for the direct delivery of retroviruses. The more invasive approach would be to intubate the patient and lavage the airway with concentrated solutions of CFTR expressing retroviruses. Stable retroviral expression requires that the provirus integrates into chromosomal DNA. This occurs most efficiently if the recipient cells are dividing. It may be necessary to stimulate regeneration of the epithelial soon after exposure to virus. This could be accomplished with mechanical or chemical irritation of the airway.

The less morbid approach would be to deliver the normal CFTR gene to airway epithelial cells *in vivo* by a nebulized preparation that can be inhaled. Many different pharmacologic agents are efficiently delivered to a large surface of the airway by nebulized treatments. It is possible that the beneficial effect achieved by this method may be transient. It may, therefore, be necessary to give repeated doses of the drug. The gene delivery system used for direct gene introduction may not have to be viral

based. Direct inhalation of DNA protein complexes or DNA expression vectors in liposomes may be a safer and more effective gene delivery system than retroviruses.

Transplantation of Genetically Modified Airway Epithelial Cells

This approach to somatic gene therapy of CF is similar in concept to bone marrow directed gene therapy. We would propose to isolate airway epithelial cells from the CF patient, establish cultures of the cells, use recombinant retroviruses described in this invention to stably correct the defect in the cells, and transplant the genetically modified cells into the patient so they can repopulate the airway. In order to achieve efficient repopulation in the airway with genetically modified cells, it may be necessary to perturb the integrity of the endogenous epithelial lining through mechanical or chemical irritation.

Alternative Gene Transfer Delivery Systems

Other gene deliveries systems for genetic correction of CF defects also fall within the scope of the present invention. For these experiments plasmid-based DNA vectors will be used. An example of such a vector is BA-CFTR BQ presented in Figure 7. This is a simple 7762 bp transfection-based vector in which transcription is initiated from actin flanking systems and terminated from heterologous 3' polyadenylation β sequences.

The vector was constructed in the following manner. The backbone contained sequences from PC18 (nucleotide 6928 to 4553) the 5' flanking region of the chicken β actin gene (nucleotide 6928 to 7754) and 3' flanking sequences of Bovine growth hormone polyadenylation signal (nucleotide 4827 to 4553). The full length CFTR sequences spanning the entire coding region, and containing the three nucleotide changes discussed earlier, were removed from the vector CFTR on a Sac I to Sal I fragment, and cloned into the vector backbone described above.

It will be appreciated by those skilled in the art that this vector could be used in several gene delivery systems.

Lipofection

The previously described procedure is based on the encapsidation of DNA liposomes. When cells are incubated with liposomes, they take up the DNA and express it. We proposed to dilute DNA of an expression vector and lipid (DOTMA) to 1.5 ml in Hepes buffered saline and mix these constituents to form lipid-DNA complexes. Liposomes could then be used to transfect airway cells *in vivo* by lavaging an intubated patient with liposome containing solution or by administering the liposomes by inhalation.

DNA-Protein Complexes

An alternative approach to targeted gene delivery is through the formation of a DNA protein complex. This type of gene transfer substrate is constructed in the following manner. A polypeptide ligand for a receptor on a respiratory epithelial cell is conjugated to polylysine with ethylidene diamino carbodiimide as described. This protein conjugate is complexed to DNA of a transfection vector by mixing equal mass quantities of protein conjugate and DNA in 0.25 molar sodium chloride. The DNA/protein complex is taken up by respiratory airway cells and the gene is expressed. This could be used to directly deliver the CFTR gene to airway epithelial cells *in vivo* using the approaches described for liposomes.

It is apparent that many modifications and variations of this invention as set forth as may be made without departing from the spirit and scope thereof. The specific embodiments described herein are given by way of example only and the invention is limited only by the terms of the appended claims.

SEQUENCE LISTING

```
                  ↓ ↓
AAT TGG AAG CAA ATG ACA TCA CAG CAG GTC AGA GAA AAA GGG TTG  45

AGC GGC AGG CAC CCA GAG TAG TAG GTC TTT GGC ATT AGG AGC TTG  90

AGC CCA GAC GGC CCT AGC AGG GAC CCC AGC GCC CGA GAG ACC ATG 135
                                                        M

CAG AGG TCG CCT CTG GAA AAG GCC AGC GTT GTC TCC AAA CTT TTT 180
 Q   R   S   P   L   E   K   A   S   V   V   S   K   L   F

TTC AGC TGG ACC AGA CCA ATT TTG AGG AAA GGA TAC AGA CAG CGC 225
 F   S   W   T   R   P   I   L   R   K   G   Y   R   Q   R

CTG GAA TTG TCA GAC ATA TAC CAA ATC CCT TCT GTT GAT TCT GCT 270
 L   E   L   S   D   I   Y   Q   I   P   S   V   D   S   A

                  *
GAC AAT CTA TCT GAA AAA TTG GAA AGA GAA TGG GAT AGA GAG CTG 315
 D   N   L   S   E   K   L   E   R   E   W   D   R   E   L

      *
GCT TCA AAG AAA AAT CCT AAA CTC ATT AAT GCC CTT CGG CGA TGT 360
 A   S   K   K   N   P   K   L   I   N   A   L   R   R   C

TTT TTC TGG AGA TTT ATG TTC TAT GGA ATC TTT TTA TAT TTA GGG 405
 F   F   W   R   F   M   F   Y   G   I   F   L   Y   L   G

GAA GTC ACC AAA GCA GTA CAG CCT CTC TTA CTG GGA AGA ATC ATA 450
 E   V   T   K   A   V   Q   P   L   L   L   G   R   I   I

GCT TCC TAT GAC CCG GAT AAC AAG GAG GAA CGC TCT ATC GCG ATT 495
 A   S   Y   D   P   D   N   K   E   E   R   S   I   A   I

TAT CTA GGC ATA GGC TTA TGC CTT CTC TTT ATT GTG AGG ACA CTG 540
 Y   L   G   I   G   L   C   L   L   F   I   V   R   T   L


CTC CTA CAC CCA GCC ATT TTT GGC CTT CAT CAC ATT GGA ATG CAG 585
 L   L   H   P   A   I   F   G   L   H   H   I   G   M   Q

                                          *
ATG AGA ATA GCT ATG TTT AGT TTG ATT TAT AAG AAG ACT TTA AAG 630
 M   R   I   A   M   F   S   L   I   Y   K   K   T   L   K

      *
CTG TCA AGC CGT GTT CTA GAT AAA ATA AGT ATT GGA CAA CTT GTT 675
 L   S   S   R   V   L   D   K   I   S   I   G   Q   L   V

AGT CTC CTT TCC AAC AAC CTG AAC AAA TTT GAT GAA GGA CTT GCA 720
 S   L   L   S   N   N   L   N   K   F   D   E   G   L   A
```

```
TTG GCA CAT TTC GTG TGG ATC GCT CCT TTG CAA GTG GCA CTC CTC  765
 L   A   H   F   V   W   I   A   P   L   Q   V   A   L   L

ATG GGG CTA ATC TGG GAG TTG TTA CAG GCG TCT GCC TTC TGT GGA  810
 M   G   L   I   W   E   L   L   Q   A   S   A   F   C   G

CTT GGT TTC CTG ATA GTC CTT GCC CTT TTT CAG GCT GGG CTA GGG  855
 L   G   F   L   I   V   L   A   L   F   Q   A   G   L   G

                                                         *
AGA ATG ATG ATG AAG TAC AGA GAT CAG AGA GCT GGG AAG ATC AGT  900
 R   M   M   M   K   Y   R   D   Q   R   A   G   K   I   S

GAA AGA CTT GTG ATT ACC TCA GAA ATG ATT GAA AAT ATC CAA TCT  945
 E   R   L   V   I   T   S   E   M   I   E   N   I   Q   S

GTT AAG GCA TAC TGC TGG GAA GAA GCA ATG GAA AAA ATG ATT GAA  990
 V   K   A   Y   C   W   E   E   A   M   E   K   M   I   E

                                     *
AAC TTA AGA CAA ACA GAA CTG AAA CTG ACT CGG AAG GCA GCC TAT 1035
 N   L   R   Q   T   E   L   K   L   T   R   K   A   A   Y

GTG AGA TAC TTC AAT AGC TCA GCC TTC TTC TTC TCA GGG TTC TTT 1080
 V   R   Y   F   N   S   S   A   F   F   F   S   G   F   F

GTG GTG TTT TTA TCT GTG CTT CCC TAT GCA CTA ATC AAA GGA ATC 1125
 V   V   F   L   S   V   L   P   Y   A   L   I   K   G   I

ATC CTC CGG AAA ATA TTC ACC ACC ATC TCA TTC TGC ATT GTT CTG 1170
 I   L   R   K   I   F   T   T   I   S   F   C   I   V   L

CGC ATG GCG GTC ACT CGG CAA TTT CCC TGG GCT GTA CAA ACA TGG 1215
 R   M   A   V   T   R   Q   F   P   W   A   V   Q   T   W

TAT GAC TCT CTT GGA GCA ATA AAC AAA ATA CAG GAT TTC TTA CAA 1260
 Y   D   S   L   G   A   I   N   K   I   Q   D   F   L   Q

AAG CAA GAA TAT AAG ACA TTG GAA TAT AAC TTA ACG ACT ACA GAA 1305
 K   Q   E   Y   K   T   L   E   Y   N   L   T   T   T   E

GTA GTG ATG GAG AAT GTA ACA GCC TTC TGG GAG GAG GGA TTT GGG 1350
 V   V   M   E   N   V   T   A   F   W   E   E   G   F   G

GAA TTA TTT GAG AAA GCA AAA CAA AAC AAT AAC AAT AGA AAA ACT 1395
 E   L   F   E   K   A   K   Q   N   N   N   N   R   K   T

 o
TCT AAT GGT GAT GAC AGC CTC TTC TTC AGT AAT TTC TCA CTT CTT 1440
 S   N   G   D   D   S   L   F   F   S   N   F   S   L   L

GGT ACT CCT GTC CTG AAA GAT ATT AAT TTC AAG ATA GAA AGA GGA 1485
 G   T   P   V   L   K   D   I   N   F   K   I   E   R   G
```

```
CAG TTG TTG GCG GTT GCT GGA TCC ACT GGA GCA GGC AAG ACT TCA  1530
 Q   L   L   A   V   A   G   S   T   G   A   G   K   T   S

CTT CTA ATG ATG ATT ATG GGA GAA CTG GAG CCT TCA GAG GGT AAA  1575
 L   L   M   M   I   M   G   E   L   E   P   S   E   G   K

            *
ATT AAG CAC AGT GGA AGA ATT TCA TTC TGT TCT CAG TTT TCC TGG  1620
 I   K   H   S   G   R   I   S   F   C   S   Q   F   S   W

            *                                  △
ATT ATG CCT GGC ACC ATT AAA GAA AAT ATC ATC TTT GGT GTT TCC  1665
 I   M   P   G   T   I   K   E   N   I   I   F   G   V   S

TAT GAT GAA TAT AGA TAC AGA AGC GTC ATC AAA GCA TGC CAA CTA  1710
 Y   D   E   Y   R   Y   R   S   V   I   K   A   C   Q   L

GAA GAG GAC ATC TCC AAG TTT GCA GAG AAA GAC AAT ATA GTT CTT  1755
 E   E   D   I   S   K   F   A   E   K   D   N   I   V   L

GGA GAA GGT GGA ATC ACA CTG AGT GGA GGT CAA CGA GCA AGA ATT  1800
 G   E   G   G   I   T   L   S   G   G   Q   R   A   R   I

TCT TTA GCA AGA GCA GTA TAC AAA GAT GCT GAT TTG TAT TTA TTA  1845
 S   L   A   R   A   V   Y   K   D   A   D   L   Y   L   L

                                           *
GAC TCT CCT TTT GGA TAC CTA GAT GTT TTA ACA GAA AAA GAA ATA  1890
 D   S   P   F   G   Y   L   D   V   L   T   E   K   E   I

TTT GAA AGC TGT GTC TGT AAA CTG ATG GCT AAC AAA ACT AGG ATT  1935
 F   E   S   C   V   C   K   L   M   A   N   K   T   R   I

TTG GTC ACT TCT AAA ATG GAA CAT TTA AAG AAA GCT GAC AAA ATA  1980
 L   V   T   S   K   M   E   H   L   K   K   A   D   K   I

                    *
TTA ATT TTG AAT GAA GGT AGC AGC TAT TTT TAT GGG ACA TTT TCA  2025
 L   I   L   N   E   G   S   S   Y   F   Y   G   T   F   S

                                       o
GAA CTC CAA AAT CTA CAG CCA GAC TTT AGC TCA AAA CTC ATG GGA  2070
 E   L   Q   N   L   Q   P   D   F   S   S   K   L   M   G

TGT GAT TCT TTC GAC CAA TTT AGT GCA GAA AGA AGA AAT TCA ATC  2115
 C   D   S   F   D   Q   F   S   A   E   R   R   N   S   I

CTA ACT GAG ACC TTA CAC CGT TTC TCA TTA GAA GGA GAT GCT CCT  2160
 L   T   E   T   L   H   R   F   S   L   E   G   D   A   P

                *                      *o
GTC TCC TGG ACA GAA ACA AAA AAA CAA TCT TTT AAA CAG ACT GGA  2205
 V   S   W   T   E   T   K   K   Q   S   F   K   Q   T   G
```

```
                                        o
GAG TTT GGG GAA AAA AGG AAG AAT TCT ATT CTC AAT CCA ATC AAC  2250
 E   F   G   E   K   R   K   N   S   I   L   N   P   I   N

 *                      o
TCT ATA CGA AAA TTT TCC ATT GTG CAA AAG ACT CCC TTA CAA ATG  2295
 S   I   R   K   F   S   I   V   Q   K   T   P   L   Q   M

AAT GGC ATC GAA GAG GAT TCT GAT GAG CCT TTA GAG AGA AGG CTG  2340
 N   G   I   E   E   D   S   D   E   P   L   E   R   R   L

 o
TCC TTA GTA CCA GAT TCT GAG CAG GGA GAG GCG ATA CTG CCT CGC  2385
 S   L   V   P   D   S   E   Q   G   E   A   I   L   P   R

ATC AGC GTG ATC AGC ACT GGC CCC ACG CTT CAG GCA CGA AGG AGG  2430
 I   S   V   I   S   T   G   P   T   L   Q   A   R   R   R

     o
CAG TCT GTC CTG AAC CTG ATG ACA CAC TCA GTT AAC CAA GGT CAG  2475
 Q   S   V   L   N   L   M   T   H   S   V   N   Q   G   Q

                         o       *   *                   o
AAC ATT CAC CGA AAG ACA ACA GCA TCC ACA CGA AAA GTG TCA CTG  2520
 N   I   H   R   K   T   T   A   S   T   R   K   V   S   L

                                                     *
GCC CCT CAG GCA AAC TTG ACT GAA CTG GAT ATA TAT TCA AGA AGG  2565
 A   P   Q   A   N   L   T   E   L   D   I   Y   S   R   R

     o
TTA TCT CAA GAA ACT GGC TTG GAA ATA AGT GAA GAA ATT AAC GAA  2610
 L   S   Q   E   T   G   L   E   I   S   E   E   I   N   E

GAA GAC TTA AAG GAG TGC CTT TTT GAT GAT ATG GAG AGC ATA CCA  2655
 E   D   L   K   E   C   L   F   D   D   M   E   S   I   P

GCA GTG ACT ACA TGG AAC ACA TAC CTT CGA TAT ATT ACT GTC CAC  2700
 A   V   T   T   W   N   T   Y   L   R   Y   I   T   V   H

AAG AGC TTA ATT TTT GTG CTA ATT TGG TGC TTA GTA ATT TTT CTG  2745
 K   S   L   I   F   V   L   I   W   C   L   V   I   F   L

GCA GAG GTG GCT GCT TCT TTG GTT GTG CTG TGG CTC CTT GGA AAC  2790
 A   E   V   A   A   S   L   V   V   L   W   L   L   G   N

                             +                       +
ACT CCT CTT CAA GAC AAA GGG AAT AGT ACT CAT AGT AGA AAT AAC  2835
 T   P   L   Q   D   K   G   N   S   T   H   S   R   N   N

AGC TAT GCA GTG ATT ATC ACC AGC ACC AGT TCG TAT TAT GTG TTT  2880
 S   Y   A   V   I   I   T   S   T   S   S   Y   Y   V   F
```

```
TAC ATT TAC GTG GGA GTA GCC GAC ACT TTG CTT GCT ATG GGA TTC  2925
 Y   I   Y   V   G   V   A   D   T   L   L   A   M   G   F

TTC AGA GGT CTA CCA CTG GTG CAT ACT CTA ATC ACA GTG TCG AAA  2970
 F   R   G   L   P   L   V   H   T   L   I   T   V   S   K

ATT TTA CAC CAC AAA ATG TTA CAT TCT GTT CTT CAA GCA CCT ATG  3015
 I   L   H   H   K   M   L   H   S   V   L   Q   A   P   M

                 *
TCA ACC CTC AAC ACG TTG AAA GCA GGT GGG ATT CTT AAT AGA TTC  3060
 S   T   L   N   T   L   K   A   G   G   I   L   N   R   F

TCC AAA GAT ATA GCA ATT TTG GAT GAC CTT CTG CCT CTT ACC ATA  3105
 S   K   D   I   A   I   L   D   D   L   L   P   L   T   I

TTT GAC TTC ATC CAG TTG TTA TTA ATT GTG ATT GGA GCT ATA GCA  3150
 F   D   F   I   Q   L   L   L   I   V   I   G   A   I   A

GTT GTC GCA GTT TTA CAA CCC TAC ATC TTT GTT GCA ACA GTG CCA  3195
 V   V   A   V   L   Q   P   Y   I   F   V   A   T   V   P

GTG ATA GTG GCT TTT ATT ATG TTG AGA GCA TAT TTC CTC CAA ACC  3240
 V   I   V   A   F   I   M   L   R   A   Y   F   L   Q   T

TCA CAG CAA CTC AAA CAA CTG GAA TCT GAA GGC AGG AGT CCA ATT  3285
 S   Q   Q   L   K   Q   L   E   S   E   G   R   S   P   I

                     *                       *
TTC ACT CAT CTT GTT ACA AGC TTA AAA GGA CTA TGG ACA CTT CGT  3330
 F   T   H   L   V   T   S   L   K   G   L   W   T   L   R

GCC TTC GGA CGG CAG CCT TAC TTT GAA ACT CTG TTC CAC AAA GCT  3375
 A   F   G   R   Q   P   Y   F   E   T   L   F   H   K   A

                                                 *
CTG AAT TTA CAT ACT GCC AAC TGG TTC TTG TAC CTG TCA ACA CTG  3420
 L   N   L   H   T   A   N   W   F   L   Y   L   S   T   L

CGC TGG TTC CAA ATG AGA ATA GAA ATG ATT TTT GTC ATC TTC TTC  3465
 R   W   F   Q   M   R   I   E   M   I   F   V   I   F   F

ATT GCT GTT ACC TTC ATT TCC ATT TTA ACA ACA GGA GAA GGA GAA  3510
 I   A   V   T   F   I   S   I   L   T   T   G   E   G   E

GGA AGA GTT GGT ATT ATC CTG ACT TTA GCC ATG AAT ATC ATG AGT  3555
 G   R   V   G   I   I   L   T   L   A   M   N   I   M   S

ACA TTG CAG TGG GCT GTA AAC TCC AGC ATA GAT GTG GAT AGC TTG  3600
 T   L   Q   W   A   V   N   S   S   I   D   V   D   S   L

ATG CGA TCT GTG AGC CGA GTC TTT AAG TTC ATT GAC ATG CCA ACA  3645
 M   R   S   V   S   R   V   F   K   F   I   D   M   P   T
```

```
                    *
GAA GGT AAA CCT ACC AAG TCA ACC AAA CCA TAC AAG AAT GGC CAA  3690
 E   G   K   P   T   K   S   T   K   P   Y   K   N   G   Q

CTC TCG AAA GTT ATG ATT ATT GAG AAT TCA CAC GTG AAG AAA GAT  3735
 L   S   K   V   M   I   I   E   N   S   H   V   K   K   D

                                *                       *
GAC ATC TGG CCC TCA GGG GGC CAA ATG ACT GTC AAA GAT CTC ACA  3780
 D   I   W   P   S   G   G   Q   M   T   V   K   D   L   T

GCA AAA TAC ACA GAA GGT GGA AAT GCC ATA TTA GAG AAC ATT TCC  3825
 A   K   Y   T   E   G   G   N   A   I   L   E   N   I   S

TTC TCA ATA AGT CCT GGC CAG AGG GTG GGC CTC TTG GGA AGA ACT  3870
 F   S   I   S   P   G   Q   R   V   G   L   L   G   R   T

    *
GGA TCA GGG AAG AGT ACT TTG TTA TCA GCT TTT TTG AGA CTA CTG  3915
 G   S   G   K   S   T   L   L   S   A   F   L   R   L   L

AAC ACT GAA GGA GAA ATC CAG ATC GAT GGT GTG TCT TGG GAT TCA  3960
 N   T   E   G   E   I   Q   I   D   G   V   S   W   D   S

ATA ACT TTG CAA CAG TGG AGG AAA GCC TTT GGA GTG ATA CCA CAG  4005
 I   T   L   Q   Q   W   R   K   A   F   G   V   I   P   Q

                        *
AAA GTA TTT ATT TTT TCT GGA ACA TTT AGA AAA AAC TTG GAT CCC  4050
 K   V   F   I   F   S   G   T   F   R   K   N   L   D   P

TAT GAA CAG TGG AGT GAT CAA GAA ATA TGG AAA GTT GCA GAT GAG  4095
 Y   E   Q   W   S   D   Q   E   I   W   K   V   A   D   E

GTT GGG CTC AGA TCT GTG ATA GAA CAG TTT CCT GGG AAG CTT GAC  4140
 V   G   L   R   S   V   I   E   Q   F   P   G   K   L   D

TTT GTC CTT GTG GAT GGG GGC TGT GTC CTA AGC CAT GGC CAC AAG  4185
 F   V   L   V   D   G   G   C   V   L   S   H   G   H   K

CAG TTG ATG TGC TTG GCT AGA TCT GTT CTC AGT AAG GCG AAG ATC  4230
 Q   L   M   C   L   A   R   S   V   L   S   K   A   K   I

TTG CTG CTT GAT GAA CCC AGT GCT CAT TTG GAT CCA GTA ACA TAC  4275
 L   L   L   D   E   P   S   A   H   L   D   P   V   T   Y

                    *
CAA ATA ATT AGA AGA ACT CTA AAA CAA GCA TTT GCT GAT TGC ACA  4320
 Q   I   I   R   R   T   L   K   Q   A   F   A   D   C   T

GTA ATT CTC TGT GAA CAC AGG ATA GAA GCA ATG CTG GAA TGC CAA  4365
 V   I   L   C   E   H   R   I   E   A   M   L   E   C   Q
```

```
CAA TTT TTG GTC ATA GAA GAG AAC AAA GTG CGG CAG TAC GAT TCC  4410
 Q   F   L   V   I   E   E   N   K   V   R   Q   Y   D   S

ATC CAG AAA CTG CTG AAC GAG AGG AGC CTC TTC CGG CAA GCC ATC  4455
 I   Q   K   L   L   N   E   R   S   L   F   R   Q   A   I

        *                                               *
AGC CCC TCC GAC AGG GTG AAG CTC TTT CCC CAC CGG AAC TCA AGC  4500
 S   P   S   D   R   V   K   L   F   P   H   R   N   S   S

AAG TGC AAG TCT AAG CCC CAG ATT GCT GCT CTG AAA GAG GAG ACA  4545
 K   C   K   S   K   P   Q   I   A   A   L   K   E   E   T

GAA GAA GAG GTG CAA GAT ACA AGG CTT TAGAGAGCAGCATAAATG       4590
 E   E   E   V   Q   D   T   R   L

TTGACATGGGACATTTGCTCATGGAATTGGAGCTCGTGGGACAGT               4635

CACCTCATGGAATTGGAGCTCGTGGAACAGTTACCTCTGCCTCAG               4680

AAAACAAGGATGAATTAAGTTTTTTTTTAAAAAAGAAACATTTGG               4725

TAAGGGGAATTGAGGACACTGATATGGGTCTTGATAAATGGCTTC               4770

CTGGCAATAGTCAAATTGTGTGAAAGGTACTTCAAATCCTTGAAG               4815

ATTTACCACTTGTGTTTTGCAAGCCAGATTTTCCTGAAAACCCTT               4860

GCCATGTGCTAGTAATTGGAAAGGCAGCTCTAAATGTCAATCAGC               4905

CTAGTTGATCAGCTTATTGTCTAGTGAAACTCGTTAATTTGTAGT               4950

GTTGGAGAAGAACTGAAATCATACTTCTTAGGGTTATGATTAAGT               4995

AATGATAACTGGAAACTTCAGCGGTTTATATAAGCTTGTATTCCT               5040

TTTTCTCTCCTCTCCCCATGATGTTTAGAAACACAACTATATTGT               5085

TTGCTAAGCATTCCAACTATCTCATTTCCAAGCAAGTATTAGAAT               5130

ACCACAGGAACCACAAGACTGCACATCAAAATATGCCCCATTCAA               5175

CATCTAGTGAGCAGTCAGGAAAGAGAACTTCCAGATCCTGGAAAT               5220

CAGGGTTAGTATTGTCCAGGTCTACCAAAAATCTCAATATTTCAG               5265

ATAATCACAATACATCCCTTACCTGGGAAAGGGCTGTTATAATCT               5310

TTCACAGGGGACAGGATGGTTCCCTTGATGAAGAAGTTGATATGC               5355

CTTTTCCCAACTCCAGAAAGTGACAAGCTCACAGACCTTTGAACT               5400

AGAGTTTAGCTGGAAAAGTATGTTAGTGCAAATTGTCACAGGACA               5445

GCCCTTCTTTCCACAGAAGCTCCAGGTAGAGGGTGTGTAAGTAGA               5490
```

TAGGCCATGGGCACTGTGGGTAGACACACATGAAGTCCAAGCATT 5535

TAGATGTATAGGTTGATGGTGGTATGTTTTCAGGCTAGATGTATG 5580

TACTTCATGCTGTCTACACTAAGAGAGAATGAGAGACACACTGAA 5625

GAAGCACCAATCATGAATTAGTTTTATATGCTTCTGTTTTATAAT 5670

TTTGTGAAGCAAAATTTTTTCTCTAGGAAATATTTATTTTAATAA 5715

TGTTTCAAACATATATTACAATGCTGTATTTTAAAAGAATGATTA 5760

TGAATTACATTTGTATAAAATAATTTTTATATTTGAAATATTGAC 5805

TTTTTATGGCACTAGTATTTTTATGAAATATTATGTTAAAACTGG 5850

GACAGGGGAGAACCTAGGGTGATATTAACCAGGGGCCATGAATCA 5895

CCTTTTGGTCTGGAGGGAAGCCTTGGGGCTGATCGAGTTGTTGCC 5940

CACAGCTGTATGATTCCCAGCCAGACACAGCCTCTTAGATGCAGT 5985

TCTGAAGAAGATGGTACCACCAGTCTGACTGTTTCCATCAAGGGT 6030

ACACTGCCTTCTCAACTCCAAACTGACTCTTAAGAAGACTGCATT 6075

ATATTTATTACTGTAAGAAAATATCACTTGTCAATAAAATCCATA 6120

CATTTGTGTAAAAAAAAAAAAAAAA

WE CLAIM:

1. A method of treatment of a defect in the gene for CFTR in a target cell comprising the steps of:

a) providing a normal CFTR gene; and

b) transferring in the normal CFTR gene into the cell under conditions favorable for transfer and complementation.

2. The method of Claim 1, wherein the step of transferring further comprises the step of viral vector-mediated gene transfer.

3. The method of Claim 1, wherein the step of transferring further comprises the step of plasmid vector-mediated gene transfer.

4. The method of Claim 1, wherein the step of transferring further comprises the step of lipofection.

5. The method of Claim 2, wherein the step of transferring further comprises the step of DNA protein complex-mediated gene transfer.

6. The method of Claim 1, wherein the target cell is an epithelial cell.

7. The method of Claim 2, wherein the virus of the step of viral-mediated gene transfer comprises a retrovirus.

8. The method of Claim 6, wherein the epithelial cell is an airway epithelial cell.

9. The method of Claim 6, wherein the epithelial cell is pancreatic cell.

10. The method of Claim 6, wherein the epithelial cell is selected from a group consisting of sweat gland, intestinal, liver and kidney cells.

11. The method of Claim 7, wherein the retrovirus is replication-defective.

12. A recombinant viral vector for treating a defect in the gene for CFTR in a target cell, the vector comprising:

a) the DNA of at least a portion of the genome of a virus which portion is capable of infecting the target cell; and

b) a normal CFTR gene operatively linked to the DNA portion of the viral genome and capable of expression in the target cell *in vivo* or *in vitro*.

13. The recombinant vector of Claim 12, wherein the virus is a retrovirus.

14. The recombinant vector of Claim 13, wherein the retroviral genome is replication-defective.

15. The recombinant vector of Claim 13, further comprising PLJ.

16. The recombinant vector of Claim 15, wherein the vector provirus is substantially as shown in Figure 1A.

17. The recombinant vector of Claim 13, wherein the defect being treated causes cystic fibrosis.

18. The recombinant vector of Claim 12, wherein the target cell is an epithelial cell.

19. The recombinant vector of Claim 17, wherein the epithelial cell is pancreatic.

20. The recombinant vector of Claim 17, wherein the epithelial cell is an airway epithelial cell.

21. The recombinant vector of Claim 17, wherein the epithelial cell is a cell selected from the group consisting of sweat gland, intestinal, liver and kidney cells.

22. The recombinant vector of Claim 18, further comprising PLJ.

23. The recombinant vector of Claim 19, further comprising PLJ.

24. The recombinant vector of Claim 15, wherein the normal CFTR gene contains silent mutations stabilizing expression.

25. A method of treating cystic fibrosis in a patient comprising the steps of:

a) providing a recombinant viral vector comprising DNA of at least a portion of a retroviral genome, which portion is capable of infecting an epithelial cell and a normal CFTR gene operatively linked thereto and capable of expression in the target cell; and

b) contacting an epithelial cell population of the patient under conditions favorable for transduction for a time sufficient for transduction to occur in at least a portion of the cell population contacted.

26. The method of Claim 25, wherein the contacting is done *in vivo*.

27. The method of Claim 25, wherein the recombinant viral vector is replication-defective.

28. The method of Claim 25, wherein the cell population comprises pancreatic cells.

29. The method of Claim 25, wherein the cell population comprises airway epithelial cells.

30. The method of Claim 25, wherein the cell population comprises a cell population selected from the group consisting of sweat gland cells, intestinal cells, liver cells and kidney cells.

31. The method of Claim 25, wherein the cell population is contacted *in vitro* and at least of portion thereof reintroduced into the patient after transduction.

32. The method of Claim 27, wherein the recombinant retroviral vector comprises PLJ.

33. A CF cell which expresses a normal CFTR gene introduced therein through retroviral transduction.

34. The cell of Claim 23, wherein the cell is derived from a PLJ-CFTR clone selected from the group consisting of PLJ-CFTR clone 1 and 3 through 20.

35. A complementation assay for a nucleic acid sequence comprising a putatively defective CFTR gene, the assay comprising:

a) a CFTR defective cell population;

b) a recombinant vector capable of infecting and transferring a CFTR gene to the CFTR defective cell population and carrying the nucleic acid sequence being assayed;

c) means for screening for the presence and/or level of CFTR function in the cell population after infection by the recombinant vector.

36. A method of assaying for putative CF mutations in a nucleic acid sequence, the method comprising the steps of:

a) providing a vector recombined with the nucleic acid sequence;

b) providing a CFTR defective cell population;

c) infecting the CFTR defective cell population with the recombined vector under conditions favorable to transduction for a time sufficient for transduction to occur; and

d) screening for complementation of CFTR function in the transduced cell population after infection.

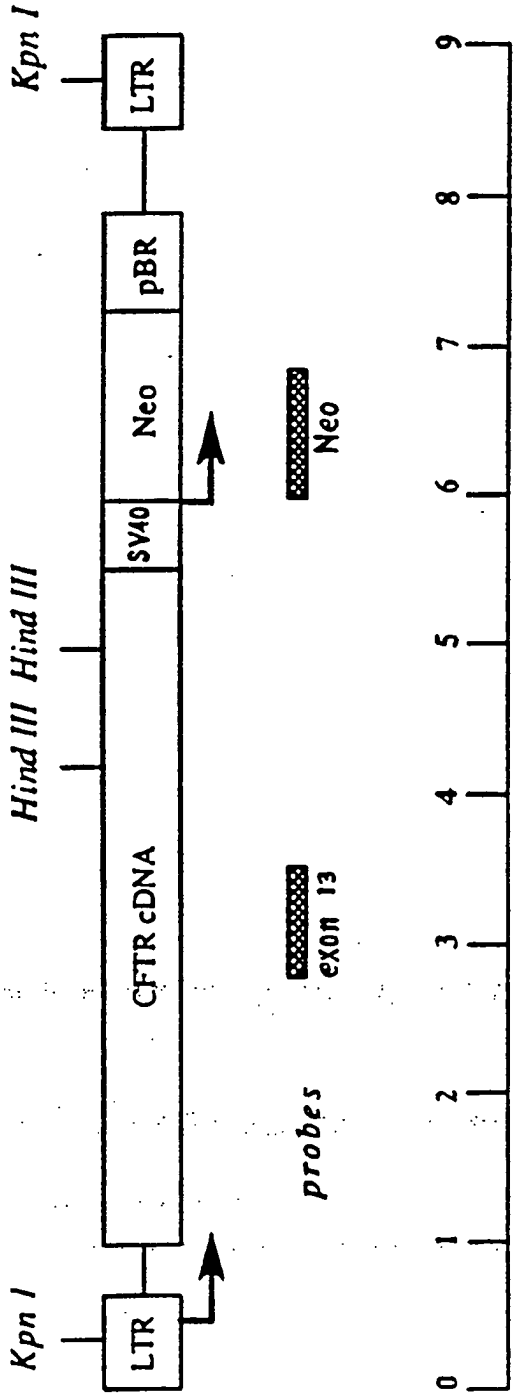

Kpn I
LTR
pBR
Neo
SV40
Hind III Hind III
CFTR cDNA
Kpn I
LTR
probes
exon 13
Neo
0
1
2
3
4
5
6
7
8
9
scale in kilobases

FIG. 1A.

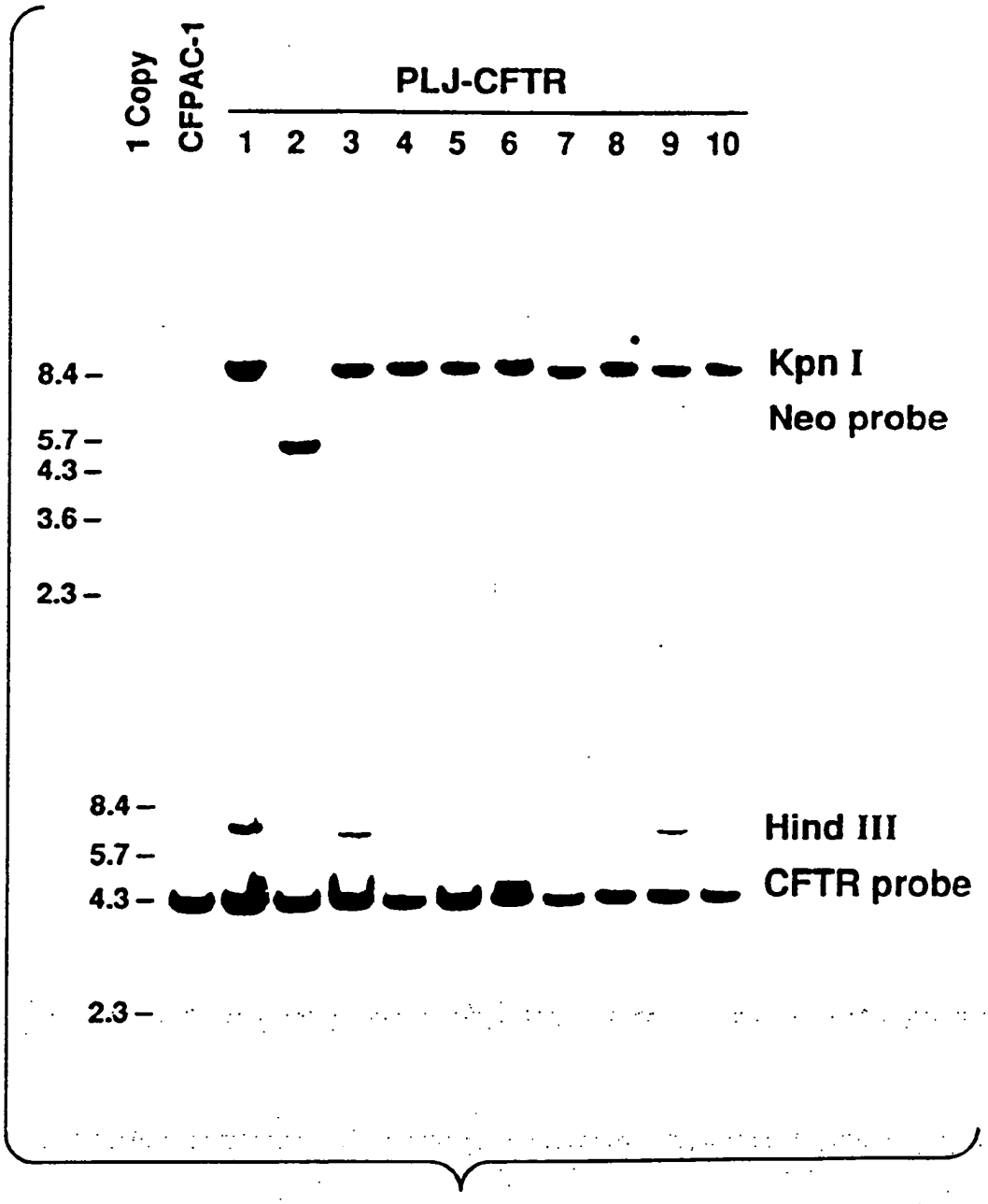

1 Copy CFPAC-1
PLJ-CFTR
1 2 3 4 5 6 7 8 9 10
8.4 –
5.7 –
4.3 –
3.6 –
2.3 –
Kpn I
Neo probe
8.4 –
5.7 –
4.3 –
2.3 –
Hind III
CFTR probe

Fig. 1B.

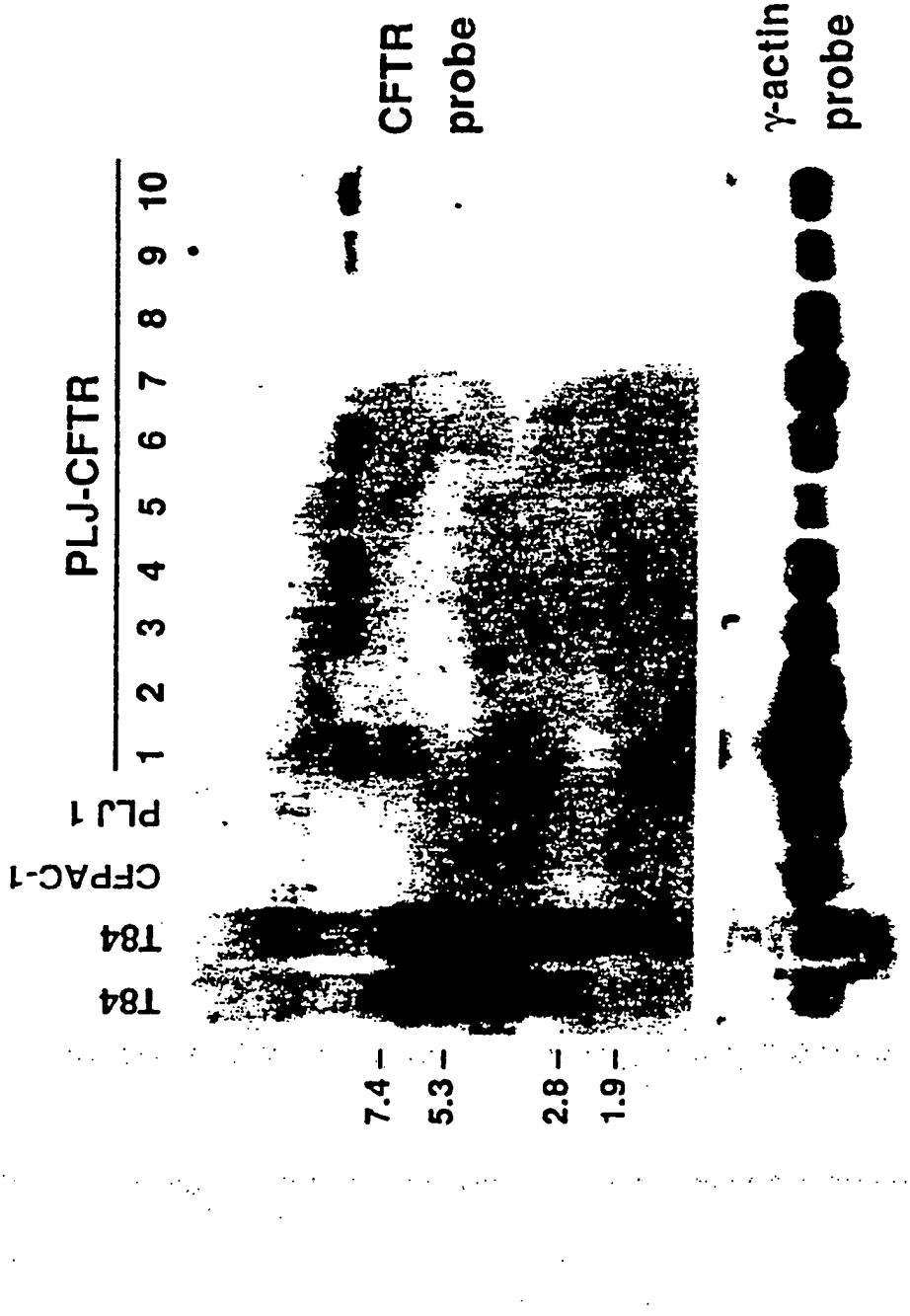

CFTR probe
γ-actin probe
PLJ-CFTR
1 2 3 4 5 6 7 8 9 10
PLJ 1
CFPAC-1
T84
T84
7.4
5.3
2.8
1.9

Fig. 2.

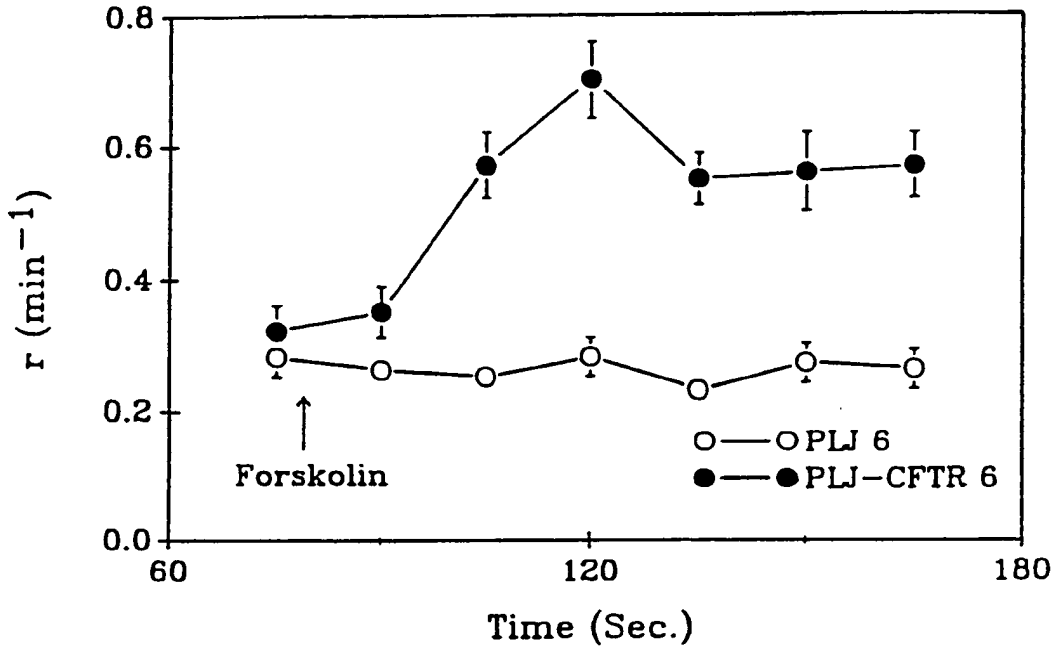

0.8
0.6
0.4
0.2
0.0
r (min^{-1})
60
120
180
Time (Sec.)
Forskolin
PLJ 6
PLJ–CFTR 6

Fig. 3A.

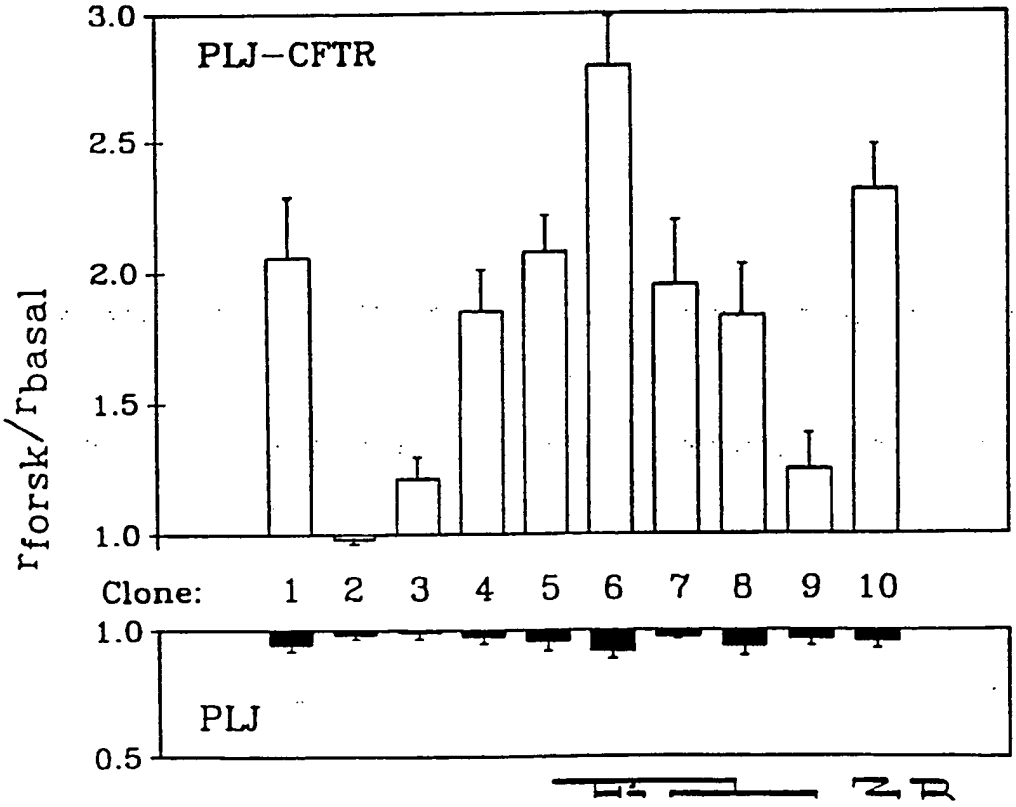

3.0
2.5
2.0
1.5
1.0
r_{forsk}/r_{basal}
PLJ–CFTR
Clone: 1 2 3 4 5 6 7 8 9 10
1.0
0.5
PLJ

Fig. 3B.

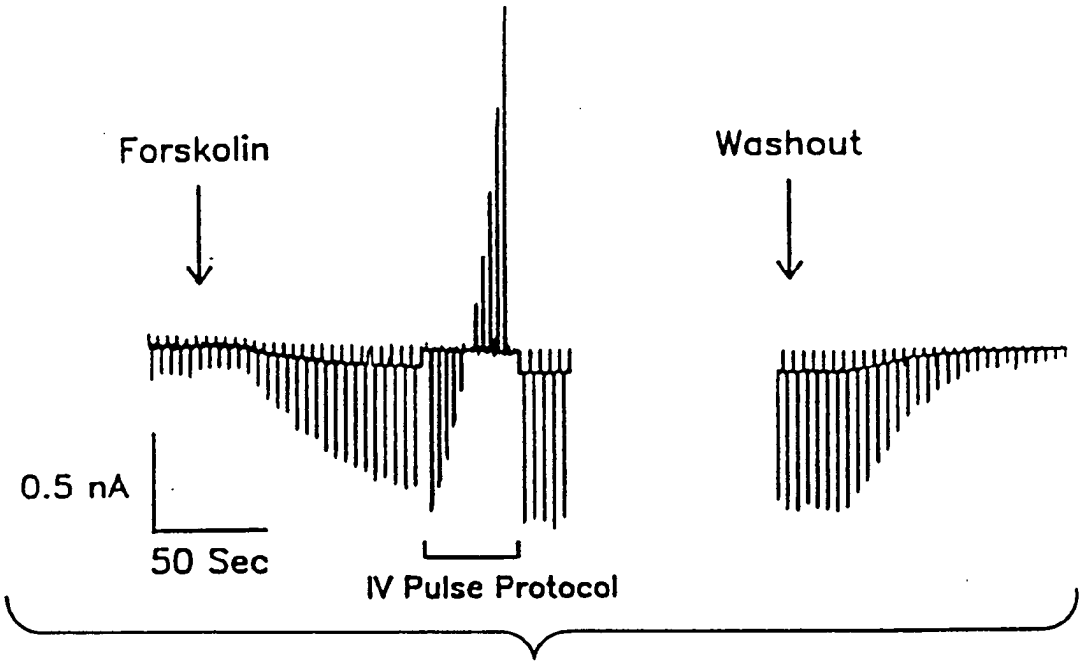

Forskolin
Washout
0.5 nA
50 Sec
IV Pulse Protocol

Fig. 4A.

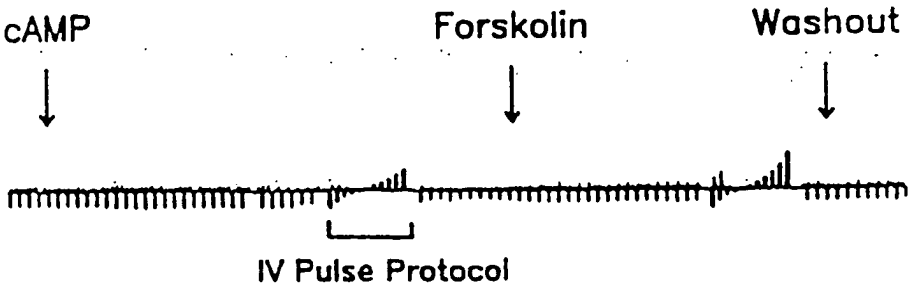

cAMP
Forskolin
Washout
IV Pulse Protocol

Fig. 4B.

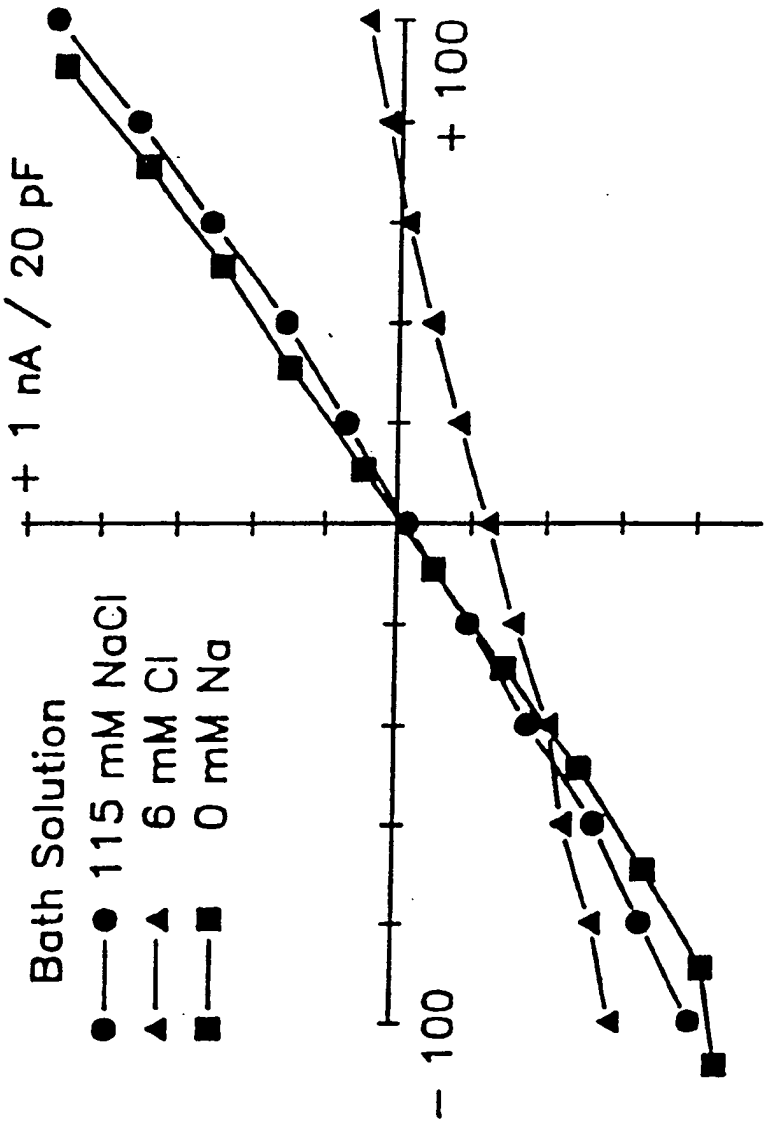
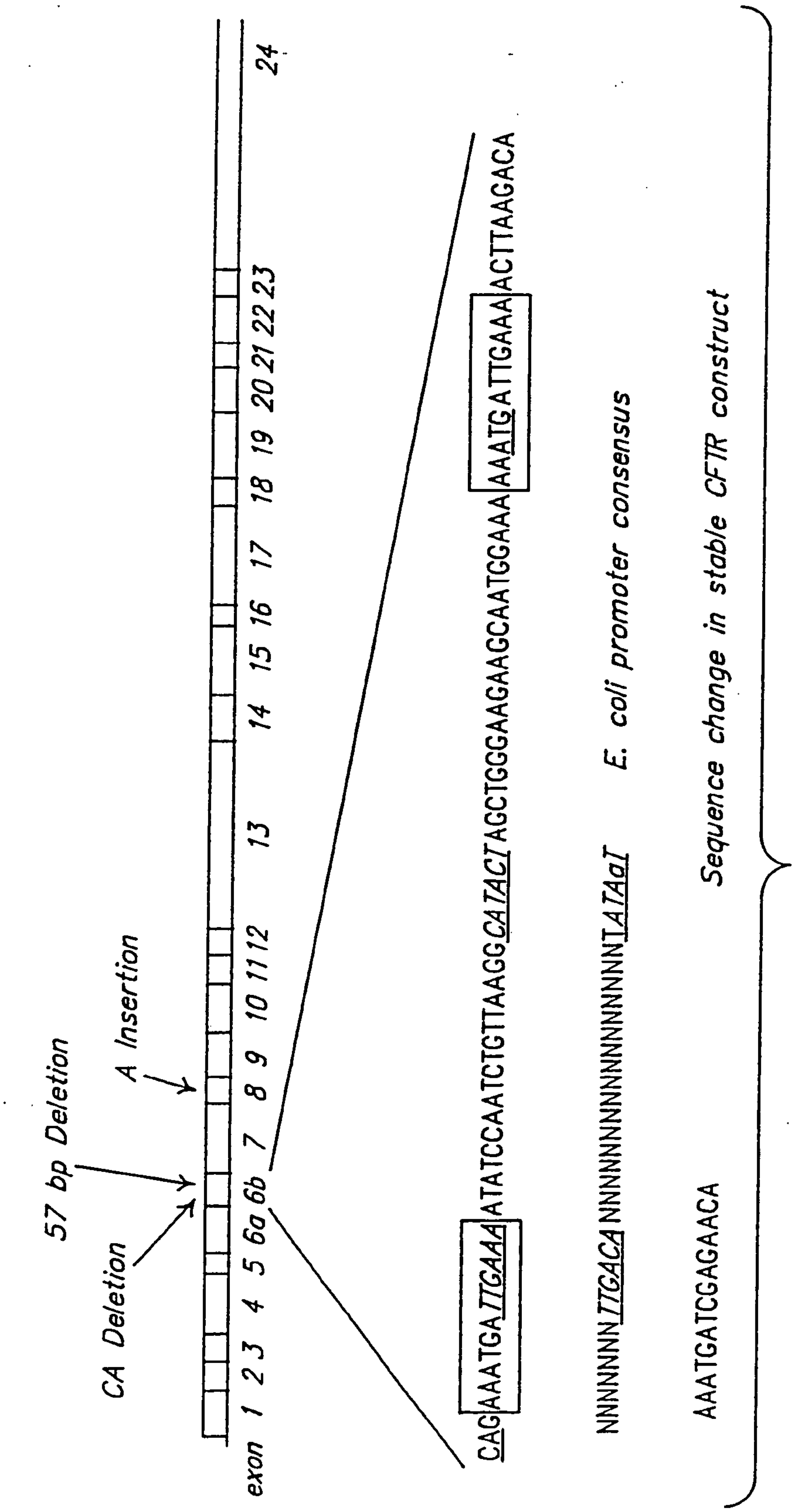
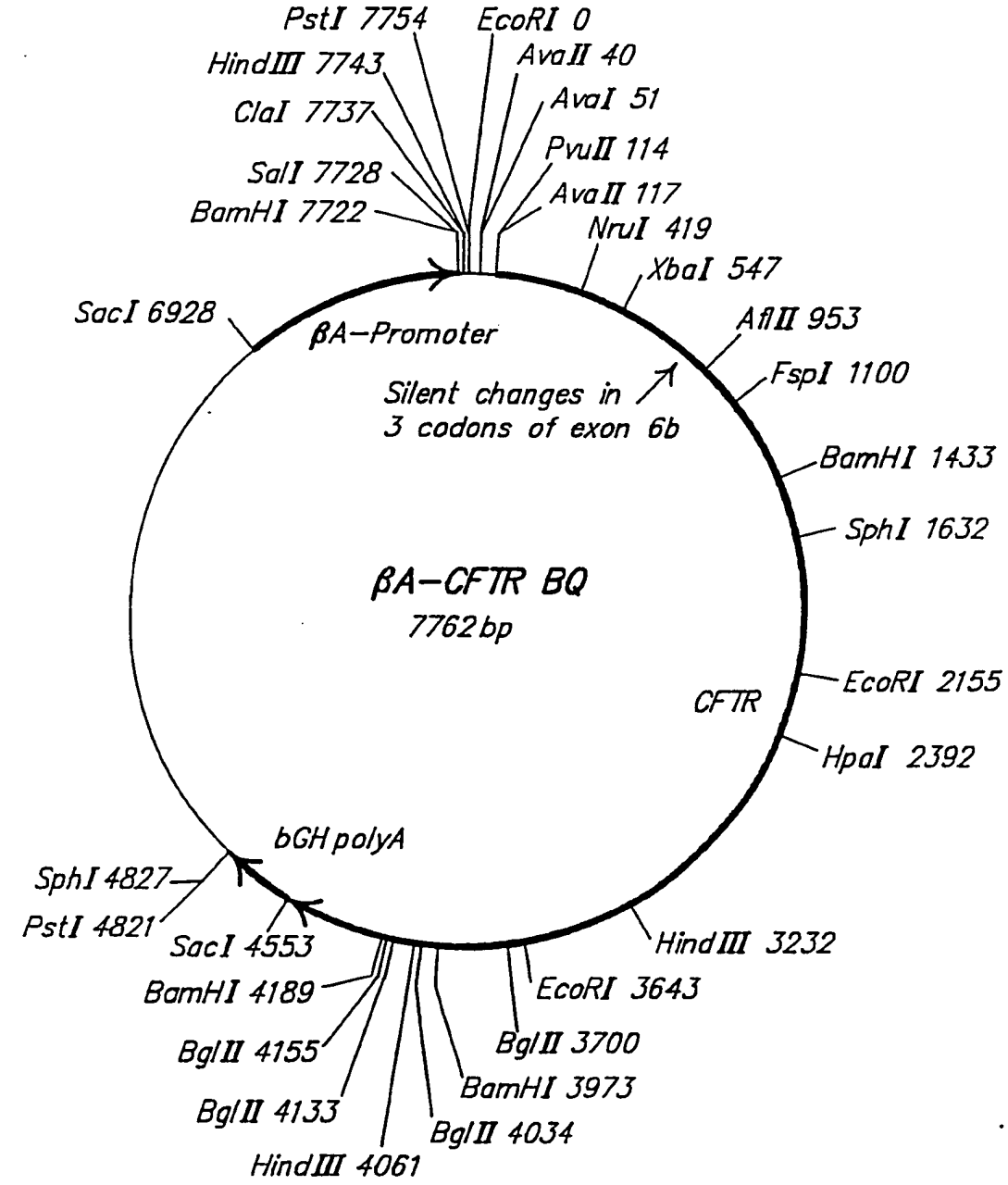


Fig. 4C


exon 1 2 3 4 5 6a 6b 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
CA Deletion
57 bp Deletion
A Insertion
CAG AAATGA TTGAAA ATATCCAATCTGTTAAGG CATACT AGCTGGAAGAGCAATGGAAA AAATG ATTGAAA ACTTAAGACA
NNNNNNN TTGACA NNNNNNNNNNNNN TATAaT
E. coli promoter consensus
AAATGATCGAGAACA
Sequence change in stable CFTR construct



Normal CFTR exon 6b sequence CAGAAATGATTGAAAATATCCAATCT...

CFTR BQ exon 6b sequence CAGAAATGATCGAGAACATCCAATCT...

FIG. 7.

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US91/06660

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) [6]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC(5): C12P 21/06; C12N 5/00, 15/00; A61K 49/00; C12Q 1/68

US.Cl.: 435/69.1, 240.26, 172.2, 173.3; 424/9; 435/6

II. FIELDS SEARCHED

Minimum Documentation Searched [7]

Classification System	Classification Symbols
U.S.Cl.	435/69.1, 172.2, 173.3, 240.26; 424/9; 435/6

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched [8]

III. DOCUMENTS CONSIDERED TO BE RELEVANT [9]

Category *	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No. [13]
Y,P	US,A, 4,980,286 (Morgan et al.) 25 December 1990, see all document.	1-11, 33-34
Y	US,A,4,868,116 (Morgan et al.) 19 September 1989, see all documents.	33-34
A,P	Science, Vol. 250, issued 05 October 1990, Green et al "Chromosomal region of the cystic fibrosis gene in yeast artificial chromosome: a model for the human genome mapping", pp. 94-98, see entire document.	1-11, 25-36
Y A	Science, Vol. 235, issued 27 February 1987, Collins et al. "Construction of a general human chromosome jumping library with application to cystic fibrosis", pp. 1047-1049, see entire document.	25-32 1-11, 33-36
A	Science, Vol. 230, issued 29 November 1985, Tsui et al. "Cystic fibrosis locus defined by a genetically linked polymorphic DNA marker", pages 1054-1057, see entire document.	1-11, 25-36

* Special categories of cited documents: [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
01 January 1992	05 FEB 1992
International Searching Authority	**Signature of Authorized Officer**
ISA/US	Miguel Escallon, Ph.D. gp

Form PCT/ISA/210 (second sheet) (Rev. 11-87)

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No
A	Genomics, Vol. 2, issued 1988, Drumm et al., "Physical mapping of the cystic fibrosis region by pulse-field gel electrophoresis", pages 346-354, see pages 350-352.	1-11, 25-36
A	Science, Vol. 233, issued August 1986, Fritzzell et al., "Altered regulation of airway epithelia cell chloride chanel in cystic fibrosis", pages 358-360, see entire document.	1-11, 25-36
A	Nature, Vol. 322, issued 31 July 1986, Welsh et al., "Chloride chanels in cystic fibrosis airway epithelia", pages 468-470, see entire document.	1-11, 25-36
A/Y	"Genetics and Epithelia Cell Dysfunction in cystic fibrosis," published 1987 by Allan R. Liss, Inc. pages 59-71, see entire document.	1-11,25-32,35,36 / 33-34
A/Y	The Journal of Cell Biology, Vol. 107, No. 6, issued 1989, Jensen et al., "Chloride chanel in culture sweat cell gland epithelia in in cystic fibrosis", see abstract 788.	1-11,25-32,35,36 / 33-34
A/Y	Proceedings of the National Academy of Science USA, Vol. 82, issued October 1983, Stuts et al., "Chloride uptake into cultured airway epithelia cells from cystic fibrosis patients and normal individuals", pages 6677-6687, see entire document.	1-11,25-32,35-36 / 33-34
A/Y	The Journal of Cell Science, Vol. 87, issued 1987, Harris et al., "Establishment of a tissue culture for epithelia cells derived from human pancreas: a model for the study of cystic fibrosis", pages 695-703, see entire document.	1-11,25-32,35,36 / 33-34
A	In Vitro Cellular and Developmental Biology, Vol. 21, No. 10, issued October 1985, Collie et al., "Culture of sweat gland epithelia cells from normal individuals and patients with cystic fibrosis", pages 597-602, see entire document.	1-11, 25-36
A	Science, Vol. 245, issued 08 September 1989, Rommens et al. "Identification of the cystic fibrosis gene: chromosome walking and jumping", pages 1059-1065, see entire document.	1-11, 25-36

FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET

V. ☐ OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE [1]

This international search report has not been established in respect of certain claims under Article 17(2) (a) for the following reasons:

1. ☐ Claim numbers . because they relate to subject matter [12] not required to be searched by this Authority, namely:

2. ☐ Claim numbers . . . . . , because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out [13], specifically:

3. ☐ Claim numbers_____, because they are dependent claims not drafted in accordance with the second and third sentences of PCT Rule 6.4(a).

VI. ☒ OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING [2]

This International Searching Authority found multiple inventions in this international application as follows:

SEE ATTACHMENT

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application.

2. ☒ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims: 1-11, 25-32, 33-36

telephone practice

3. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:

4. ☐ As all searchable claims could be searched without effort justifying an additional fee, the International Searching Authority did not invite payment of any additional fee.

Remark on Protest

☐ The additional search fees were accompanied by applicant's protest.

☐ No protest accompanied the payment of additional search fees.

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C12N 15/12, 1/21, 5/16
C07K 13/00

A1

(11) International Publication Number: WO 92/05252

(43) International Publication Date: 2 April 1992 (02.04.92)

(21) International Application Number: PCT/CA91/00341

(22) International Filing Date: 20 September 1991 (20.09.91)

(30) Priority data:
9020632.7 21 September 1990 (21.09.90) GB

(71) Applicant *(for all designated States except US)*: HSC RESEARCH AND DEVELOPMENT LIMITED PARTNERSHIP [CA/CA]; 88 Elm Street, Toronto, Ontario M5G 1X8 (CA).

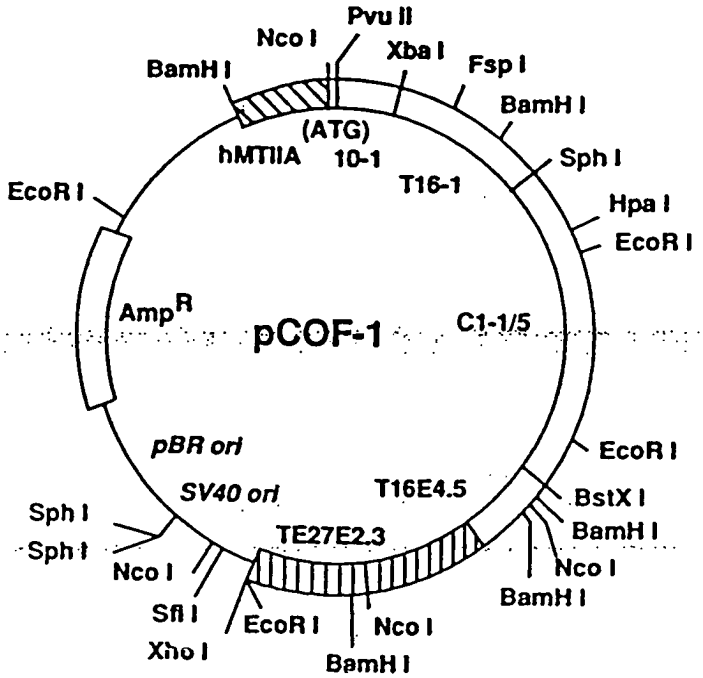
(72) Inventors; and
(75) Inventors/Applicants *(for US only)* : TSUI, Lap-Chee [CA/CA]; 94 Willowbridge Road, Toronto, Ontario M9R 3Z4 (CA). ROMMENS, Johanna, M. [CA/CA]; 199 Bogert Avenue, Willowdale, Ontario M2N 1L1 (CA).

(74) Agent: WOODLEY, John, H.; Sim & McBurney, 330 University Avenue, Toronto, Ontario M5G 1R7 (CA).

(81) Designated States: AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FI, FR (European patent), GB (European patent), GR (European patent), IT (European patent), LU (European patent), NL (European patent), NO, SE (European patent), US.

Published
With international search report.

(54) Title: STABLE PROPAGATION OF MODIFIED FULL LENGTH CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR PROTEIN cDNA IN HETEROLOGOUS SYSTEMS



(57) Abstract

A modified DNA sequence encoding full length cystic fibrosis transmembrane conductance regulator protein is provided to facilitate propagation and/or expression of the protein in living cells and in particular, bacterial cells. The modified DNA sequence comprises at least one of the 13 base pair repeat of exon 6b of the normal gene encoding the conductance regulator protein, as one or more normal nucleotides of the 13 base pair repeat substituted with an alternate nucleotide which however continues to code for the corresponding normal amino acid. Mammalian cells transfected with a vector containing the modified DNA sequence enhances chlorine conductance through the cell wall.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland			SN	Senegal
CI	Côte d'Ivoire	KR	Republic of Korea	SU+	Soviet Union
CM	Cameroon	LI	Liechtenstein	TD	Chad
CS	Czechoslovakia	LK	Sri Lanka	TG	Togo
DE*	Germany	LU	Luxembourg	US	United States of America
DK	Denmark	MC	Monaco		

+ Any designation of "SU" has effect in the Russian Federation. It is not yet known whether any such designation has effect in other States of the former Soviet Union.

STABLE PROPAGATION OF MODIFIED FULL LENGTH CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR PROTEIN cDNA IN HETEROLOGOUS SYSTEMS

FIELD OF THE INVENTION

This invention relates to modifications in the cDNA of full length cystic fibrosis transmembrane conductance regulator protein (CFTR) which facilitates propagation and/or expression in heterologous systems.

BACKGROUND OF THE INVENTION

Cystic fibrosis (CF) is the most common, life-threatening, autosomal recessive disease in the Caucasian population. Approximately 1 in 2,500 live births is affected by this genetic disorder. Obstructive lung disease, pancreatic enzyme insufficiency and elevated sweat electrolytes are the hallmarks for CF but the severity of these symptoms vary from patient to patient. Patients with CF usually die at an early age due to lung infection. With recent advances in clinical treatments, which are directed against the symptoms, the mean survival age for patients has increased to 26 years.

Despite intensive research efforts for the past fifty years, the basic defect in CF remains to be speculative. It is generally believed that the heavy mucus found in the respiratory tracts and the blockage of exocrine secretion from the pancreas are due to imbalance in water secretion which is the consequence of a defect in the regulation of ion transport in the epithelial cells.

The precise localization of the CF locus on the long arm of chromosome 7, region q31, facilitated the recent isolation of the responsible gene. The CF gene spans 250 kilobase pairs (kb) of DNA and encode a mRNA of about 6,500 nucleotides in length. The CFTR gene is disclosed and claimed in United States application S.N. 401,609 filed August 31, 1989. That application is co-owned by the applicant of this application.

Expression of this gene could be observed in a variety of tissues that are affected in CF patients, for example, lung, pancreas, liver, sweat gland and nasal epithelia. An open reading frame spanning 1480 amino acids could be deduced from the overlapping cDNA clones isolated. The putative protein as noted is called "Cystic Fibrosis Transmembrane Conductance Regulator" or CFTR for short, to reflect its possible role in the cells. The predicted molecular mass of CFTR is about 170,000.

Based on sequence alignment with other proteins of known functions, CFTR is thought to be a membrane-spanning protein which can function as a cyclic AMP-regulated chloride channel. The internal sequence identity between the first and second half of CFTR resembles the other prokaryotic and eukaryotic transport proteins, most notably, the mammalian P-glycoprotein.

The most frequent mutant allele of the CF gene involves a three base pair (bp) deletion which results in the deletion of a single amino acid residue (phenylalanine) at position 508, within the first ATP-binding domain of the predicted polypeptide. Although this mutation (ΔF508) accounts for about 70% of all CF chromosomes, there is marked difference in its proportion among different populations. The remaining 30% of mutations in the CF gene appear to be heterogeneous and most of them are rare, with some represented by only single examples, as referenced in applicant's Canadian patent application filed July 9, 1990.

The mutation screening study confirms that the ATP-binding domains detected by sequence alignment is important for CFTR function as multiple, different mutations have been found for many of the highly conserved amino acid residues in these regions. The locations of the various mutations also identified other functionally important regions in CFTR. There is, for example, a section three bp deletion resulting in the

omission of an isoleucine residue at position 506 or 507 of the putative protein. While amino acid substitutions at these positions are apparently not disease-causing, this observation argues that the length of the peptide is more critical than the actual amino acid residue in the 506-508 region. Further, the existence of a large number of nonsense, frameshift as well as mRNA splicing mutations in the CF gene implies that absence of CFTR is not incompatible with life.

The varied symptoms among different CF patients suggest that disease severity is at least in part related to the mutations in the CF gene. Such association, which is expected to be concordant among patients within the same family, as they should have the same genotype at the CF locus, is observed for pancreatic function. Approximately 85% of CF patients are severely deficient in pancreatic enzyme secretion, thus diagnosed as pancreatic insufficient (PI), and the other 15% have sufficient enzyme, thus pancreatic sufficient (PS). Family studies showed that there was almost complete concordance of the pancreatic status among patients within the same family, leading to the suggestion that PI and PS are predisposed by the patients' genotypes. Subsequent studies showed that patients homozygous for the ΔF508 mutation were almost exclusively PI. This information may be useful in disease prognosis.

There are other mutations that would be classified in the same group as ΔF508, the so-called severe mutant alleles with respect to pancreatic function. In contrast, patients with one or two copies of other class (i.e. mild) of alleles are expected to be PS. Meconium ileus which is observed in about 30% of CF patients appears to be a clinical variation of PI and not directly determined by the CF genotype. Other clinical manifestations are more complicated and no apparent association has yet been detected.

With the identification of the CF gene, a better understanding of the basic defect and pathophysiology of the disease can now be attained. Progress and advance are being made in studies of the regulatory mechanisms governing the expression of this gene, and of the biosynthesis and subcellular localization of the protein (through generation of antibodies against various parts of the protein). In addition, it is important to develop effective assay systems for the function of CFTR. This information may be useful in development of rational therapies, including gene therapy.

In order to obtain a DNA sequence containing the entire coding region of CFTR, it is necessary to construct a full-length cDNA from overlapping clones previously isolated. A major difficulty has been encountered in the process, however. As the various proportions of the full-length cDNA is being linked together by standard procedures; i.e., restriction enzyme cutting and ligation, with plasmid vector in *Escherichia coli*, frequent sequence rearrangement has been detected in the resulting construct.

For purposes of better understanding of the regulatory functions of the CFTR protein and also for purposes of gene and drug therapy, it is useful to be able, in a commercial way, to propagate and express the normal CFTR gene and various mutant CFTR genes in a variety of hosts which include bacteria, yeast, molds, plant and animal cells and the like.

Although propagation and expression of the cDNA sequence for the CFTR gene can be achieved in some vehicles, there are, however, the aforementioned difficulties in obtaining stable propagation of the cDNA in some types of bacteria, particularly *E. coli*. It is thought that the cDNA contains sequence portions which, when propagated in the bacteria, results in a toxic effect which is countered by lack of propagation of the cDNA in the microorganism.

INVENTION SUMMARY

We have discovered that a modification of a cDNA repeat sequence in exon 6 of the CFTR gene without modifying the amino acids encoded by the changed codons facilitates propagation and/or expression of the CFTR protein in living cells and in particular bacterial cells.

According to an aspect of the present invention, a modified DNA sequence derived from a gene coding for cystic fibrosis transmembrane conductance regulator (CFTR) protein, the gene having at least 27 exons of which normal cDNA codes for CFTR protein, the normal cDNA including exons 6a and 6b wherein exon 6b includes a 13-bp repeat,

the modified DNA sequence comprises at least one of the 13 bp repeats of exon 6b having one or more normal nucleotides of the 13 bp repeat substituted with an alternate nucleotide which continues to code for a corresponding normal amino acid.

According to another aspect of the present invention, a DNA construct for use in a recombinant vector comprises the modified cDNA.

According to a another aspect of the present invention, a vector which comprises the DNA construct and a promoter sequence for the DNA construct.

According to futher aspect of the present invention, a host cell for producing CFTR protein, the host cell comprises the above vector whereby expression of the vector in the host cell produces CFTR protein.

According to another aspect of the present invention, a CFTR protein isolated and purified from culture of the host cell.

According to a further aspect of the present invention, a mammalian cell transfected with the above vector to enhance Cl^- conductance through a cell wall.

DESCRIPTION OF DRAWINGS

Figure 1 is the nucleotide sequence of the CF gene and the amino acid sequence of the CFTR protein;

Figure 2 is a schematic diagram of the CF gene and its protein product. (A) Gene structure with exons represented by open boxes; (B) Computer-predicted primary structure of CFTR; and (C) Model of CFTR showing possible relationship with cellular membrane and potential glycosylation site (-CHO);

Figure 3 is a DNA sequence of exon 6b. The nucleotide positions (877-1002) for exon 6 correspond to the previous cDNA numbering scheme (Riordan et al 1989). The 13 bp direct repeats are underlined. The consensus prokaryotic transcription signals (at positions -35 and -10) are included for comparison. The modified nucleotides in pCOF-1; pBQ6.2 and pBQ4.7 are shown underneath.

Figure 4 is a schematic diagram showing the three basic steps in the construction of full-length CFTR cDNA.

Figure 5 is the generation of the 5' end of CFTR cDNA for the expression vector pCOF-1. Two overlapping, oppositely oriented oligonucleotides, Cof1 (44-mer) and Cof2 (40-mer), were used in the generation of the exon 1 sequence. A double stranded molecule was synthesized from these two oligonucleotides with the Klenow fragment of DNA polymerase. An NcoI site was introduced at the initiation codon to facilitate subsequent cloning. The PCR-generated fragment was digested with NcoI and PvuII and ligated to the remaining portions of the CFTR cDNA, including the PvuII to XbaI fragment of cDNA clone 10-1, the PCR-modified middle fragment and the 3' fragment . The procedures used to generate the two latter segments are described with reference to Figures 6 and 7. The position of the human metallothionein IIA promoter, the bacterial ampicillin-resistance gene, the pBR origin of replication and the SV40 origin of replication are marked.

Figure 6 is 3'end construction. The 3' coding region of the expression vector (pCOF-1) and pBQ6.2 was constructed in two sequential cloning steps. First, the 3' most portion, including the end of the coding sequence and the 3' untranslated region, was prepared by ligating a fragment from T16-4.5 (generated by BcoI partial digestion) to a fragment from the genomic clone TE27 (Rommens et al 1989). The resulting fragment was then joined with the SphI-BstXI fragment from C1-1/5 (riordan et al 1989) at the BsXI restriction site to yield a 4 kb 3' end fragment.

Figure 7 is a schematic diagram showing the synthesis of the middle segment containing the modified sequence. Details are described herein.

Figure 8 is propagation vectors pBQ4.7 and pBQ6.2. The construction of these plasmids are provided in description of the invention. Key restriction sites are marked. The sizes of the transcripts generated by the T7 RNA polymerase (from the PstI site to each of the marked restriction sites) are also indicated (in kilobases).

Figure 9 is DNA sequencing ladder showing the modified sequence in exon 6b. The sequencing reaction was performed with the use of an oligonucleotide primer (19-mer) corresponding to a sequence in exon 6a starting at nucleotide 771 (5'-GCTAATCTGGGAGTTGTT-3'). The altered nucleotide in the sequence are underlined.

Figure 10 is <u>in vitro</u> translation of CFTR. The products from <u>in vitro</u> translation reactions were separated on a 10% polyacrylamide -SDS gel. The gel was stained by Coomassie Blue after electrophoresis; it was then soaked in Enhance® (New England Nuclear) and dried at 60°C under vacuum before exposure to X-Omat film (Kodak). Lane 1: no added RNA; lane 2: protein translated from the Brome Mosaic Virus RBA (as size and reaction conditions control); lanes 3 and 4: duplicate samples of capped mRNA prepared by T7 RNA polymerase on template pQB6.2 digested with HpaI. The positions for

the protein molecular weight standards (purchased from BRL Labs) are indicated on the left in kilodaltons. The expected product of82.6 kb (from the RNA generated from the pBQ6.2 plasmid) is marked. Also marked (by arrows) are the sizes of the BMC protein products.

Figure 11 is DNA hybridization analysis of integrated plasmids. Genomic DNA samples were prepared from each cell line, digested with BamHI and separated on a 0.7% agarose gel in Tris-acetate buffer. The DNA lanes were transferred to Zetabind (Bio-Rad Labs) membrane. The autoradiograph shows the hybridization result with the CFTR cDNA probe. Lanes 1 and 2 contain human and mouse genomic DNA, respectively. Lanes 3, 4, 25 and 26 contain non-transfected mouse LTK-cell DNA. Lanes 5 to 9 contain DNA from HAT-resistance clones obtained by transfection with pSTK7 only. Lanes 10 to 18 contain DNA from HAT resistant clones obtained by co-transfection with pSTK7 and pCOF-1 and lanes 19 to 24 with pSTK7 and pCONZ. Molecular weight markers are shown on the left (in kb). The two diagnostic bands are 2.3 kb (the 5'end of the CFTR gene plus the expression vector, promoter sequence) and 2.5 kb (CFTR internal segment); they are indicated on the right by the arrows.

Figure 12 is RNA hybridization analysis. Total RNA was extracted from each cell line, purified through CsCl gradient centrifugation, and separated on a 1.0% formaldehyde agarose gel in MOPS-acetate buffer. The samples were transferred to Zetabind (Bio-Rad Labs) membrane and hybridized sequentially with radioactively labeled CFTR cDNA (panel A) and the TK cDNA probe (panel B). The samples in lanes 1 and 21 were prepared from the colonic cell line T84; lane 2 from fresh mouse salivary tissue; and lane 3 from untransfected LTK-cells. Lanes 4 to 7 contain RNA prepared from independent clones generated from pSTK7 transfection; lanes 8 to 15 from cotransfection with pSTK7 and pCOF-1; and lanes 16 to 20 from co-transfection with pSTK7 and pCONZ. The 28S and

18S rRNA bands are indicated. The position anticipated for the full-length transcript (6.2 kb) is also marked. The expected transcript for the HSV YK gene is 2.2 kb.

Figure 13 is Western Blot analysis. Total protein extracts were prepared from the indicated cell lines, separated on a 7.5% polyacrylamide-SDS gel and transferred to nitrocellulose as described herein. Lane 1: extract from line 2a-4A; lane 2: 4a-2D; lane 3:4a-3I; lane 4:6b-J. The blot was incubated with a mouse monoclonal antibody and followed by anti-mouse antibody. The immune complexes were visualized with alkaline phosphatase as described herein. Protein molecular weight standards are shown on the left. The arrow on the right marks the 170 kilodalton band.

Figure 14 is expression vector pCOF-1. The complete CFTR coding region (open boxes) is positioned downstream from the human metallothionein IIa (hMTIIa) promoter (hatched box). The human metallothionein IIa initiation codon is joined with that of CFTR at an Nco I site introduced by the synthetic oligonucleotides. 5'-CACTGCAGACCATGGAGAGGTCGCCTCTGGAAAAGGCCAGCGTT-3' and 5'-GACTGCAGCTGAAAAAAAGTTTGGAGACAACGCTGGCCTT-3'. The DNA sequence from exons 2 to 24 and its 3' flanking region was derived from clones 10-1 (Pvu II-Xba I), T16-1 (Xba I-Sph I), C1-1/5 (Sph I-BstX1), T16E4.5 (BstXI-Nco I), and TE27E2.3 (Nco I-Xho I), as indicated. In pCOFΔF508, the 1-kilobase (kb) BamIII-Hpa a fragment was replaced with the corresponding fragment from clone C1-1/5.

Figure 15 is RNA and protein analysis of mouse L cell lines expressing human CFTR. (A) Total RNA from Caco2 (5μg, lane 1), LTK⁻ (10 μg, lane 2), 6B-I (10 μg, lane 3), 4a-3I (10 μg, lane 4), and 5-2D (10 μg, lane 5) cell lines were electrophoresed on a 1% formaldehyde/agarose gel, transferred to Hybond-N (Amersham), an hybridized with ^{32}P-labelled cDNA probe. The 6.2-kb CFTR mRNA of the Caco2 cell line is indicated with the arrow. The relative positions of the 28S and

18S rRNAs as indicated (B) Protein fractions from nuclei and mitochondria (lanes A) and curde light membranes (lanes B) from the cell lines T84, LTK$^-$, 6B-I, 4a-3I and 5-2D. Bands were visualized by ^{125}I-labelled rabbit anti-mouse antibody. Molecular mass in kDa are indicated.

Figure 16 is video microscopic detection of Cl^- permeability in single L cells, SPQ fluorescence intensities (F) are expressed reltive to SPQ fluorescence intensity in the absence of Cl^- quenching (F_o; the direction of changes in F_o/F reflect parallel changes in intracellular Cl^- concentration. (A-C) Single cells, each indicated by a different symbol, are shown containing a frameshift construct leading to a predicted truncated CFTR (A); a mutant construct CFTR$_\Delta$F508 (B); and an intact construct, CFTR (C). Forskolin (10 μM) and NO^-_3 medium were perfushed over the cells for the periods indicated by solid bars. Time scale in C applies to A-C. (D) A single CFTR cell was repeatedly pulsed with NO^-_3 medium. Rate of fluorescence chage as well as peak response during constant pulses (1 min) was unchanged over the time course of exposure to forskolin, indicating that the cAMP-induced Cl^- permeability is sustained. (E) Simultaneous determinations of Cl^- concentrations (F_o/F) (open and solid squares) and cell volume, expressed as relative area of constant optical section (open and solid circles), in a single cell expressing CFTR (solid squares and circles)) and a single cell expressing truncated CFTR (open squares and circles). As indicated, the cells were exposed to 5 μM gramicidin and to 10 μM forskolin.

Figure 17 is whole-cell Cl^- currents in transfected L cell. (A) Time course of whole-cell currents measured from cells transfected with the frame-shift CFTR control construct (solid diamonds) and the intact CFTR construct (open squares) after the addition of a solution containing 10 μM forskolin, 1 mM isobutylmethylxanthine, and 100 μM N^6, $O^{2'}$-dibutylyladenosine 3',5'-cyclic monophosphate. The arrow indicates the time of solution

addition. Outward currents were measured at E_m = +20 mV. (B) Whole cell current-voltage relationships for a CFTR-expressing cell before (Left) and 3 min after (Right) induced activation. The current scale for the nonstimulated cells is shown enlarged as indicated. The dashed line indicates the zero-current level. (C) Mean current-voltage relationships for CFTR cells (n = 10) before and after activation (solid circles and open squares, respectively). Replacement of bath NaCl with sodium gluconate resulted in a shift in reversal protention of cAMP-activated currents (open triangle) (n = 4).

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The normal CFTR gene sequence and protein sequence is shown in Figure 1. The sequence consists of cDNA identified and isolated from chromosome of mammalian cells in accordance with the procedure identified in co-applicants United States patent application SN 401,609. It is believed that the cDNA of CFTR is made up of at least 27 exons as identified in applicant's pending Canadian patent application filed July 9, 1990.

The exact cause of the sequence instability is unknown, but the region frequently involved in this rearrangement is a 13 bp direct repeat in exon 6a of the CFTR gene as shown in Figure 3. This region also shows sequence identity with consensus regulatory sequences in E. coli as noted in Figure 3. It is possible that transcription initiates from this sequence, resulting in a product which is toxic to the bacterial host. It is also possible that the DNA sequence itself elicits an inhibitory response. In consequence, the transformation frequency and growth of the desired plasmid construct become extremely poor. The rearrangement of this sequence detected in the recovered plasmids may have alleviated its toxicity in bacteria. Thus, the inability to obtain a full-length CFTR cDNA in the original cloning experiment (Rommens et al, 1989) may also be explained.

Since most of the spontaneous sequence rearrangements detected disrupts the open reading frame of the full-length cDNA, none of the resulting clones are expected to be useful in functional complementation studies. Although it is possible to use alternative host/vector systems to overcome this problem, the convenience of the E. coli/pBR322-based cloning system required a better system.

The region of interest with respect to enhancing propagation of the cDNA in microorganisms and in particular in E. coli is in exon 6. A 13 base pair repeat has been found in exon 6. With reference to Figure 1, the region of exon 6 is identified. More specifically, within exon 6 the repeat sequences occur at positions 923 through 935 and at 981 through 993. These repeat sequences are separated by approximately 44 base pairs.

It has been discovered that by modifying one or more of the base pairs in region 924 through 936 or 982 through 994 without changing the amino acid encoded by the respective codon results quite surprisingly in the propagation of the cDNA in a suitable microorganism, such as E. coli. The modified cDNA, in one of the repeat base pairs sections, does not produce a toxic effect in the microorganism so that a stable propagation of the cDNA can proceed. The benefit in the stable propagation of cDNA is not only from the standpoint of expression of the gene to produce the protein, but also for producing multiple copies of the cDNA for the CFTR gene to enable other uses of the gene, such as in drug and gene therapy for correcting the effects of CF.

Construction of Full Length cDNA

Two types of full length cDNA clones have been constructed to evaluate the modified cDNA. One of them is for efficient propagation in E. coli and the other one for expression in mammalian cells.

Since the entire coding region of CFTR is present in multiple overlapping clones, it is necessary to construct

the full-length cDNA in three separate steps (Figure 4): construction of the 5'end of the cDNA (from the beginning of the coding region to the XbaI site in exon 5), (Rommens et al, 1989) generation of the middle segment (from the XBaI site to the SphI site in exon 10) including the critical region by the polymerase chain reaction (PCR: Saiki et al, 1985), and (Riordan et al, 1989) construction of the 3'end of the cDNA from different existing clones (from SphI site to exon 24). The starting clones of this work have been described previously (Riordan et al, 1989).

To construct the 5'end of the cDNA in the mammalian expression vector, oligonucleotides were used to synthesize the open reading frame between the initiation codon and the end of exon 1 (the PvuII site). This fragment was inserted between the promotor of vector pSG3X and exon 2 of cDNA clone 10-1 (see Figure 5). A single nucleotide modification (C to G) was introduced at the position immediately after the initiation codon to facilitate subsequent manipulation. The alteration also introduced an amino acid change from Glutamine to Glutamate.

To construct the 3' end of the cDNA, DNA fragments were obtained from three existing clones (Figure 6): SphI to BstXI (exons 10-20) from C1-1/5 (Riordan et al, 1989), BstXI to NcoI (exons 20-24) from T16-4.5 (Riordan et al, 1989), and NcoI to EcoRI (exon 24 to about 100 bp downstream from the polyadenylation signal) from TE27 (Rommens et al, 1989). The genomic DNA fragment included at the end of this construct to ensure proper transcription termination and mRNA processing. This 3'end construct was used for both the expression and propagation vectors.

The cDNA clone T16-1 (Riordan et al, 1989) was used as the template for generation of the middle segment, which was used for both vector systems. To introduce a sequence modification in one of the 13 bp repeat region

of exon 6, the in vitro mutagenesis procedure based oligonucleotide-directed PCR (Higuchi et al, 1988; Ho et al, 1989) was used. Briefly, two overlapping segments, spanning exons 3 to 6 (536 bp) and exon 6 to 17 (867 bp), were generated from T16-1 by PCR with two sets of oligonucleosides (see Figure 7). The two overlapping segments were joined by the PCR procedure with the outermost, flanking oligonucleotide primers. The product, 1.38 kb in size, was then digested with XbaI and SphI to generate the modified middle segment.

The pBluescript® vector from Stratagene was used as the basis propagation vector. The 5' end of the full-length cDNA in this vector was derived from cDNA clone 10-1 (between the PstI and XbaI sites). The resulting clones were named pBQ4.7 and pBQ6.2 (Figure 8).

The plasmid pSGM3X was used as the basic mammalian expression vector. It is similar to p SGM1 previously reported by Meakin et al (1987), except that the human metallothionein promoter was inserted in an opposite orientation and that a XhoI site was inserted in the KpnI site within the Ecogpt gene. The resulting expression vector was named pCOF-1 (Figure 5).

Sequence Verification

Since errors are often introduced into PCR products due to infidelity of the Taq polymerase, the middle segment of the cDNA clones were first verified by DNA sequencing. The procedure of dideoxy termination was used according to the US Biochemical Sequenase® kit. Consequently, it was necessary to replace a small segment (the XbaI site in exon 6 to the FspI site in exon 7) of the clone with the corresponding segment from T1601. Overlapping DNA sequencing was then performed for the entire open reading frame of each of the completed full-length cDNA constructs.

In vitro Translation

To produce large quantities of RNA from the propagation vector (pBQ4.7 and pBQ6.2) for the purpose of

in vitro translation, the mCAP® mRNA capping kit from Stratagene was used.

Since the open reading frame predicts a protein of 170,999 kilodaltons (kd), it would be difficult to produce a product of this size in vitro. It was therefore necessary to use differently prepared templates to examine various segments of the open reading frame. To prepare these templates, the propagation vector pBQ6.2) was digested with FspI, EcoRI, HpaI and XhoI, accordingly.

in vitro translation was performed with a rabbit reticulocyte lysate kit from Promega Corporation and ^{14}C-labelled leucine from Amersham.

Transfection Studies in Mouse LTK Cells

The calcium phosphate co-precipitation procedure for introduction of plasmid DNA into mouse LTK-cells has been described previously (Meakin et al, 1987). The plasmid pSTK7 was used for cotransfection with the expression vector pCOF-1. As a control vector, a plasmid construct similar to pCOF-1, except for the deletion of a single base pair in exon 1 was used; this deletion was expected to result in premature termination of translation.

The mouse LTK-cells were passaged in α-MEM medium supplemented with glutamine, 10% fetal bovine serum and antibiotics. Biochemical selection for TK positive cells was achieved in medium containing hypoxanthine, aminopterin and thymidine (HAT medium).

DNA and RNA Analyses

Plasmid and DNA samples were prepared from the bacterial cells and genomic DNA from transfected mouse L cells. Total RNA was extracted from mammalian cells for examination of gene expression. Standard procedures, essentially as described by Sambrook et al (1989), for DNA and RNA analyses were used.

Protein Analysis

Animal cells were harvested in TEN buffer (40 mM Tris-HCl, pH 7.5, 1 mM EDTA, 150 mM), after the cells

were washed in phosphate-buffered saline. Cells were collected by scraping, resuspended in 2560 mM Tris-HCl (pH 8), and stored in -80°C ready for total protein extraction. For SDS-polyacrylamide gel electrophoresis, the frozen cells were resuspended in the loading buffer and boiled for 5 min prior to electrophoresis according to the procedure of Laemmli et al (1970).

Transfer of protein from polyacrylamide gel to nitrocellulose was performed overnight at 0.4 mA according to the procedure of Towbin et al (1979). The membrane was baked at 80°C for 90 min. Detection of CFTR was accomplished with the use of a monoclonal antibody (gift of N. Kartner and J. Riordan) prepared against a portion of the predicted protein (amino acid residues 347-698) and the ProtoBlot® AP System according to procedures recommended by the supplier (Promega Corporation) with 1% BAS as a blocking agent.

Vector Construction

Three plasmid constructs were made in this study - pBQ4.7; pBQ6.2 and pCOF-1 as described in the previous section. pBQ4.7 is a full-length cDNA clone contained in the Bluescript® vector; it contains the entire coding region except the 3' untranslated region. pBQ6.2 contains the same sequence as in pBW4.7 plus the 3' untranslated region. Both of these plasmids contain modification in the first 13 bp direct repeat in exon 6b (Figure 3). The full-length cDNA construct in pCOF-1 is inserted downstream from the human metallothionein IIA promoter in the vector pSGM3X and the coding region is flanked by CFTR genomic DNA sequence at the 3' end. In addition to the modification introduced into the exon 6b region of the cDNA, as in the two previous constructs, pCOF-1 has another modification (C to G) in the first nucleotide after the presumptive initiation (ATG) codon in exon 1.

As the first step in examining their integrity, restriction enzyme digestions were used to derive a map for each of these full-length cDNA constructs. The

results showed that all of them produced a set of DNA fragments as expected. It is, however, of interest to note that there was one DNA fragment in each of the digestions which appeared to have a different mobility when compared to the pattern of the corresponding parental cDNA clones displayed on polyacrylamide gel. The mapping data located the fragments with the altered mobility to those containing the 13 bp direct repeats.

To ensure the DNA sequence was intact as designed, the middle segment containing the modified segment was examined by DNA sequencing. The result showed that 3 bp were altered in the region. While two of the modifications (T to C at position 930 and A to G at 933) were inserted as expected, an additional change (T to C at position 936) was found (Figure 9). This unanticipated alteration was found in all five clones examined, suggesting that it was introduced by an error in the synthesis of the oligonucleotide primer. Nevertheless, this error was located outside of the "-35" prokaryotic consensus sequence (Hawley and McClure 1983) and did not alter the encoded amino acid (Figure 3).

Two other undesired nucleotide changes were detected in the middle cDNA segment. Since these latter substitutions, probably introduced during PCR, would change the encoded amino acids, it was necessary to replace the region containing these changes with a segment from the original plasmid.

The DNA sequence in all three resulting clones appeared to be stable after long term propagation in *E. coli*. Their entire coding region was sequenced and no additional changes were detected. It was noted that the transformation efficiency of these constructs in bacteria increased substantially (100 times higher than the unmodified cDNA) and that the copy number of plasmids (as reflected by the yield in DNA preparation) also increased.

Protein Synthesis by in vitro Translation

To ensure that the open reading in the full-length cDNA was uninterrupted, the in vitro translation method was performed. Accordingly, the propagation vector was linearized at an appropriate restriction site in the circular plasmid and used as the template for production of RNA suitable for translation in vitro. Since the entire CFTR protein would be probably too large to be translated in vitro, the experiment also included shorter transcripts produced from templates interrupted within the coding region.

An example of the result is shown in Figure 10. The template used for this experiment was pBQ6.2 linearized at the HpaI site; the in vitro translation product was expected to be 86.2 kd. As shown in Figure 10, a band migrating at the position of the expected molecular weight is clearly visible, indicating that translation initiates at the ATG codon as predicted and continues through the HpaI site in exon 13. The identity of this protein is also confirmed by its ability to react with a monoclonal antibody against CFTR. The other, smaller proteins observed in the products are probably results of premature termination or intenal initiation of translation.

Similar experiments were performed with the plasmid treated with XbaI and EcoRI. Protein bands corresponding to the expected translation products 19 and 77.5 kd, respectively, were readily detectable on the polyacrylamide gel. These results, therefore, provided further confirmation that the predicted open reading frame was correct and intact in the full-length cDNA constructs.

Expression in Mouse L Cells

To examine if CFTR could be produced in heterologous mammalian cells, the pCOF-1 plasmid was used to transfect mouse LTK-cells. Another plasmid containing the herpes simplex virus TK gene was included in the transfection to

allow biochemical selection of cells that were successfully transfected. A plasmid (pCONZ) which suffered a single base pair deletion immediately adjacent to the 3' initiation codon was used in a parallel transfection experiment as a negative control (as the frameshift would result in no CFTR products).

DNA Analysis

HAT-resistant L cell colonies were isolated and expanded into individual mass cultures and genomic DNA were isolated from these cultures for characterization of integrated plasmid DNA. As expected, all of the HAT-resistant L cell lines was found to contain an intact HSV TK gene (as demonstrated by gel-blot hybridization analysis). In addition, the copy number for the integrated plasmid DNA was found to vary among different lines, an anticipated from the calcium phosphate co-precipitation protocol. Gel-blot hybridization analysis was then performed for the cell cultures transfected with pCOF-1 and pCONZ. The full-length cDNA clone (a 6.2 kb PstI fragment from pBQ6.2) was used as proof to examine the CFTR sequence.

Among the 15 pCOF-1/pSTK7 co-transfected cell lines, 14 were found to contain at least a portion of the CFTR cDNA, as judged by the hybridization pattern of EcoRI-digested DNA from these lines. Five of these lines shows a 1.5 kb and 2.5 kb band, as predicated for an intact cDNA. For the remaining lines, the size of the EcoRI fragment was different from those predicted, indicating rearrangement or integration occurred through this region of the cDNA. Upon further analysis with BamHI and NcoI digestion, however, only four of the five lines showed intact 5' and 3' ends of the cDNA (see Figure 11). The apparently intact clones were 4a-2C, 4a-3I, 4a-3K, and 4a-4S.

Similar DNA analysis was performed for cell lines co-transfected with pCONZ and pSTK7. Among the 10 clones examined, nine appeared to contain CFTR sequence and

eight of them showed the predicted 1.5 kb and 2.5 kb EcorI fragments (see Figure 12). Further analysis with BamHI and NcoI on five of them revealed that four (6a-1D; 6a-2F; 6b-J and 6b-K) contained intact 5' and 3' ends.

RNA Analysis

Expression of CFTR in the transformed mouse cell lines was next examined by RNA blot hybridization analysis. Two of the pCOF-1 transfected lines (4a-3I and 4a-3K) were found to express high levels of RNA (see Figure 12a) but the remaining two lines 4a-1C and 4a-4Q) were low. In addition to the 6.2 kb band expected for correctly initiated and terminated CFTR RNA in these lines; however, hybridizing sequences were also detected at the 9-10 kb range, suggesting improper initiation or termination. A similar, high expression pattern was observed for three of the lines transfected with pCONZ (6a-1A; 6a-1D and 6b-J) (Figure 12a). For clone 6a-2F, only a low level of the 6.2 kb species was detected. No RNA could be found for clone 6b-K, despite the presence of an apparent intact CFTR sequence.

The level of CFTR-derived RNA present varied among the cell lines mentioned above, with the strongest signals seen in clones 4a-3I and 6b-J. The variation was not due to poor sample preparation, as confirmed by hybridization analysis of the same blot with the TK probe (Figure 12B).

Protein Analysis

Total protein was then extracted from several representative cell lines and examined for the presence of CFTR-related products. The interpretation of the results of this experiment was complicated by the presence of cross-reactive materials present in the mouse L cell, although no CFTR-related RNA was detectable. Figure 13 is an illustration in drawing format of the Western blot results. These results are based on a monoclonal antibody (L12B4) prepared against CFTR which is able to detect faint bands (at regions 170, 90, and 74

kd) in all cell extracts. Nevertheless, the intensity of the 170 kd band, the one expected for the CFTR protein, was clearly much stronger in 4a-3I, the cell line containing an intact full-length CFTR cDNA. Only background bands were detected for 2a-4A (containing pSTK7 only), 4a-2D (incomplete pCOF-1 sequence), and 6b-J (pCONZ). The presence of the CFTR protein in 4a-3I was also confirmed in partially purified membrane fractions.

1) A single 170kdalton band was observed in the extract from the cell line 4a-3I. This cell had been demonstrated by DNA analysis to contain multiple copies of the intact promoter of the expression vector and of the CFTR gene. RNA analysis also indicated expression of high levels of mRNA for this clone.

2) In contrast, the cell line 6b-J, generated with the mutant gene does not express a reacting 170kdalton band. This line generated by co-transfection with the pCONZ plasmid was shown to have an intact promoter and gene portion. In addition, high levels of RNA were detected. Identical results for these samples were obtained even when the extracts were not subjected to heating prior to gel loading.

3) 170kdalton bands were not observed in the cell line 4a-2D generated by the transfection experiment with the pCOF-1 plasmid and the cell line 2a-4A generated by the transfection with the pTK7 plasmid only. Genomic DNA analysis of the cell line 4a-2D indicated that only portions of the CFTR gene had been incorporated. As expected, no CFTR RNA could be detected.

In conclusion, the cell line 4a-3I appears to be producing a 170kdalton CFTR protein as expected from the open reading frame of the pCOF-1 plasmid. For comparison, also shown on the protein analysis figure are crude membrane preparations of the cell lines Panc-1 (a negative control) and T84 (provided by N. Kartner and J. Riordan). Multiple bands are observed in the T84 preparation. These differences suggest that post-

translational modification including glycosylation and/or phosphorylations have occurred to the CFTR protein in T84 cells.

A CAMP-inducible chloride permeability has been detected in mouse fibroblast (L cell) lines upon stable integration of a full-length cDNA encoding the human cystic fibrosis transmembrane conductance regulator (CFTR). As indicated by a Cl^--indicator dye, the Cl^- permeability of the plasma membrane increases by 10-to 30-fold within 2 min after treatment of the cells with forskolin, an activator of adenylyl cyclase. The properties of the conductance are similar to those described in secretory epithelial cells; the whole-cell current-voltage relationship is linear and there is no evidence of voltage-dependent inactivation or activation. In contrast, this cAMP-dependent Cl^- flux is undetectable in the untransected cells or cells harboring defective cDNA constructs, including one with a phenylalanine deletion at amino acid position 508 (ΔF508), the most common mutation causing cystic fibrosis. Those observations are consistent with the hypothesis that the CFTR is a cAMP dependent Cl^- channel. The availability of a heterologous (nonepithelial) cell type expressing the CFTR offers an excellent system to understand the basic mechanisms underlying this CFTR-associated ion permeability and to study the structure and function of the CFTR.

We describe the construction and use of a mammalian expression vector to produce human CFTR in a long-term mouse fibroblast culture. We show that expression of CFTR induces a cAMP-dependent Cl^- conductance, which is normally not observed in these cells. This expression system is suitable for study of the Cl^- conductance pathway and its regulation and to provide medical treatment for CF.

Plasmid Vectors

The mammalian expression vector pCOF-1 is a derivative of pSGM3X, which is similar to pSMG1 (Meakin et al), except that the human metallothionein IIa promoter (Karin et al, 1982) was inserted in the opposite orientation and a Xho I site was inserted in the Kpn I site within the Ecogpt gene. To reconstruct the full-length CFTR cDNA in pCOF-1 (Fig. 14), the bulk of the coding region (exons 2-24) was obtained from partial cDNA clones (Riordan et al, 1989), except that the three silent nucleotide substitutions (T → C at position 930, A → G at position 933, and T → C at position 936) were introduced into the exon 6b region with oligonucleotide-mediated mutagenesis by the polymerase chain reaction (19, 20). The 3' untranslated region of CFTR in pCOF-1 was derived from the genomic DNA clone TE27E2.3 (Rommens et al, 1989). The entire coding region of exon 1 (from the initiation codon to the Pvu II site) was generated by two complementary synthetic oligonucleotides and the Klenow fragment of DNA polymerase I, where a single nucleotide substitution (C → G) was introduced immediately after the initiation codon (underlined in the legend to Fig. 14) to create a Nco I site for ligation to the human metallothionein IIa promoter. The latter substitution changed the encoded amino acid glutamine to glutamic acid. The construction of control plasmid pCONZ was similar to that of pCOF-1, except that a single nucleotide was deleted 35 base pairs downstream from the initiation codon. A truncated protein would be predicted from this frameshift construct. The plasmid pCOFΔF508 was generated by replacing sequences of exons 9-13 in pCOF-1 with the corresponding fragment from Cl-1/5, a cDNA containing the ΔF508 mutation. The full-length cDNA clone pBQ6.2 contained a 6.2 kb Pst I fragment in pBluescript (Stratagene) and was constructed similarly to pCOF-1 except that the exon 1 region was derived from clone 10-1. The integrity of the CFTR cDNA inserts in

pBQ6.2, pCOF-1, and the critical regions in the other plasmid constructs were verified by DNA sequencing. The plasmid vector (pSTK7) containing the herpes simplex virus thymidine kinase gene, used in the cotransfection experiments, has been described (Meakin et al). Bacterial cell cultures and plasmid DNA samples were prepared according to standard procedures (Sambrook et al, 1989).

Cell Culture and DNA Transfection

Each of the three test plasmids (20 μg), pCOF-1, pCONZ, and pCOFΔF508, was cotransfected with pSTK7 (1 μg) into mouse LTK⁻ cells by calcium phosphate coprecipitation (Meakin et al). Biochemical selection for thymidine kinase-positive cell was achieved in minimal essential medium supplemented with hypoxanthine/aminopterin/thymidine (HAT;gibco/brl). In some experiments, the test plasmids were linearized at the unique Sfi I site (Fig. 14). High molecular weight DNA was isolated from each clonal cell line (Miller et al, 1988), digested with restriction enzymes EcoRI, BamHI, and Nco I, and analyzed by agarose-gel-blot hybridization (Sambrook et al, 1989) with the full-length cDNA (insert from pBQ6.2) as probe. Total RNA was extracted (MacDonald et al, 1987) and analyzed by agarose-gel-blot hybridization (Sambrook et al, 1989).

Protein Analysis

Cells were homogenized in a hypotonic buffer containing 10 mM KCl, 1.5 mM $MgCl_2$, and 10 mM Tris·HCl (pH 7.4). Nucleic and mitochodria were collected by centrifugation at 4000 x g for 5 min (fraction A). A crude membrane fraction was then collected by centrifugation at 9000 x g for 15 min (fraction B). Membrane pellets were dissolved in loading buffer and separated on a $NaDodSO_4$/polyacrylamide (6%) gel (Laemmli, 1970). Proteins were transferred to nitrocellulose as described (Towbin et al, 1979) and immunodetected with monoclonal antibody M3A7 (Kartner et al, 1991).

6-Methoxy-1-(3-sulfonatopropyl)quinolinium (SPQ) Fluorescence Assay

L cells grown on glass coverslips for 1-2 days were uniformly loaded with the Cl^- -indicator dye SPQ by incubation in hypotonic (1:2 dilution) medium containing 20 mM SPQ at room temperature for 4 min. The mounted coverslip was perfused continuously at room temperature with medium containing 138 mM NaCl, 2.4 mM K_2HPO_4, 0.8 mM KH_2PO_4, 10 mM Hepes, 1 mM $CaCl_2$, 10 mM glucose, and 10 μM bumetanide (pH 7.4) on the stage of an inverted microscope. NO^-_3 medium was identical except that NO^-_3 replaces all but 10 mM Cl^-. To minimize Cl^- fluxes through nonconductive pathways, we performed the experiments in the absence of HCO^-_3 and in the presence of bumetanine, inhibiting the anion exchanger and Cl^-/cation cotransporters, respectively. Fluorescence and the differential interference contrast imagine were performed simultaneously (25-27). SPQ fluorescence intensities (F) were normalized to total SPQ fluorescence F_o, determined as F measured in the absence of intracellular Cl^-, since autofluorescence was negligible, Calibration (n = 7 cells) were performed essentially as described (Foskett, 1990). The effective quenching constant K_{Cl} was 15 M^{-1}; the testing intracellular Cl^- concentration was $\approx$ 70 mM. Cell volume changes were obtained by planimetry of differential interference contrast images (26, 27) and are presented as relative changes in the areas of the measured optical sections. By exposure of the cells to media of various osmolarities, we observed that differential interference contrast imaging of a single optical section can detect volume changes.

Whole-Cell Current Recordings

Membrane currents were measured at room temperature 12-24 hr after plating cells with whole-cell patch-clamp techniques (Hamill et al, 1981). The patch pipet contained 110 mM sodium gluconate, 25 mM NaCl, 8 mM $MgCl_2$, 10 mM hepes, 4 mM Na_2ATP, and 5 mM Na_2EGTA (pH 7.2). The

bath contained 135 mM NaCl, 2.4 mM K_2HPO_4, 0.8 mM DH_2PO_4, 3 mM $MgCl_2$, 1 mM $CaCl_2$, 10 mM Hepes, and 10 mM glucose (pH 7.2). To examine the time course of a cAMP-evoked conductance changes, membrane potentials were alternately clamped at -30 and +20 mV for 600 and 400 ms, respectively. Current-voltage relationships were determined by measuring the currents at the end of 400 - ms voltage steps from 0 mV to ± 70 mV (10 mV increments). Cell capacitance was compensated using cancellation circuitry of the EPC7 patch-clamp amplifier.

DNA Transfection

Stable mouse fibroblast cell lines containing full-length CFTR or mutant cDNA, as well as cell lines with the pSTK7 plasmid alone, were established. Four cell lines with pCOF-1 (4a-2C, 4a-3I, 4a-3K, and 4a-S) contained an intact human metallothionein IIa promoter and the CFTR coding region, as judged by DNA analysis (data not shown). Similarly, four cell lines (6a-1D, 6a-2F, 6b-I, and 6b-K) for pCONZ control cDNA with the frame-shift mutation three for pCOF$_\Delta$F508 (5-2C, 5-1A, and 5-2D) cDNA with the major CF mutation and two for pSTK7 (2a-4A and 2a-3C) were identified).

RNA Analysis

Abundant levels of RNA transcripts of ≈ 10 kb in size were detected in cDNA transected cells Fig. 15A. Although their size was larger than the anticipated 6.2 kb, the result appeared to be consistent among all cell lines. It seemed probable that an alternative polyadenylylation site(s) instead of those contained in the cDNA constructs were utilized. Thymidine kinase-specific transcripts were detected in all HAT -resistant cell lines tested.

Protein Analysis

Cells expressing CFTR mRNA (as represented by the cell line 4a-3I) contained an antibody-reacting protein band that was indistinguishable from mature CFTR expressed endogenously in membranes of the colonic

carcinoma cell line T-84 (Fig. 15B). The amount of protein was within the range of that observed for T-84 cells, with a significant portion in the light membrane fraction. Reacting bands were not detected in untransfected LTK$^-$ cells or in cell transfected with CFTR gene predicted to produce a truncated product (line 6B-I). The latter result was expected as the antibody was directed against the C terminus of the protein. No immunoreactive material was observed for CFTRΔF508 in the examined fractions (line 5-2D) (Cheng et al, 1990). It was therefore uncertain if the mutated protein was produced in these transfected cells.

SPQ Fluorescence Assay

To investigate if expression of CFTR affected Cl^- conductance, a single-cell assay based on quantitative fluorescence intensity measurements of the Cl^- indicator dye SPQ was performed. The basic protocol involved exposing cells to NO^-_3 medium followed by return to normal Cl^- medium. Since NO^-_3 is generally permeable through Cl^- channels but, unlike Cl^-, does not quench spq, CHANGES IN spq fluorescence intensities due to these anion substitutions measure unidirectional Cl^- fluxes; the rate of change measures cell Cl^- permeability. 10 μM forskolin was added subsequently to increase intracellular levels of cAMP. After a 2-min exposure to forskolin, the medium was again switched to the NO^-_3 medium, in the continued presence of forskolin. Thus each cell was used as its own control to evaluate the Cl^- permeability induced by cAMP.

Only slow changes in SPQ fluorescence intensity were observed in cells with the frame-shift DFTR construct (as represented) by line 6B-I) upon exposure to NO^-_3 medium (Fig. 16A), indicating that these cells maintained a low resting Cl^- permeability. Exposure of these transfected control cells to forskolin did not have any effect (n = 41 cells from four passages). Similar results were obtained from cells transfected with the

thymidine kinase gene only (line 2a-4A, n = 22 cells) and from untransfected cells (n = 18 cells).

Exposure of cells expressing CFTR to NO^-_3 medium similarly elicited little or no change in SPQ fluorescence intensity (as represented by line 4a-3I in Fig. 15C), indicating that CFTR expression per se did not enhance Cl^- permeability. In contrast, a second exposure to NO^-_3, during forskolin stimulation, caused a rapid loss of intracellular Cl^- (Fig. 16C). This response was highly reproducible; there was a 20- to 30-fold increase of unidirectional Cl^- flux (Illsley et al, 1987) from ≈ 0.03 mM/s to 0.9 mM/s for each of the 10 cells in the microscopic field (Fig. 16C). Of 106 cells examined from eight passages, all responded similarly. Similar results were obtained from another cell line expressing CFTR (clone 4a-3K, n = 23 cells). In the continued presence of forskolin (Fig. 16D), Cl^- permeability remained enhances at near maximal levels for as long as 30 min (n = 12 cells).

There were no volume changes during exposure of forskolin-stimulated CFTR cells to NO_{-3}, indicating that the substantial changes in the intracellular concentration of Cl^- were not associated with changes in cell salt content. Therefore, the Cl^- fluxes were likely to be associated with equal fluxes of NO^-_3 in the opposite direction. The lack of changes in intracellular Cl^- concentration or cell volume during forskolin stimulation in Cl^- medium demonstrates that neither CFTR expression nor cAMP conferred enhanced cation conductance. Together with the enhanced Cl^- permeability, cell shrinkage or swelling would have been observed if there were a high K^+ conductance or high Na^+ conductance, respectively.

Cl^- permeability was also examined in cells containing a CFTRΔF508 construct. These cells (as represented by clone 5-2D, n = 15 cells) exhibited low Cl^- permeability under resting conditions and the

permeability could not be enhanced by forskolin (Fig. 16B).

To establish that the cAMP-induced Cl^- permeability in the CFTR-expressing cells was due to activation of a Cl^- conductance, gramicidin was included in the normal Cl^- medium to increase cation conductance of the plasma membrane. Under these conditions, the presence of Cl^- channel would result in a substantial influx of both Na^+ and Cl^-, causing extensive cell swelling. Exposure to gramicidin had no effect on SPQ fluorescence or cell volume in cells with the frame-shift CFTR construct (n = 13 cells) or an intact CFTR construct (n = 33 cells) (Fig. 16E), supporting our conclusion (above) that resting Cl^- conductance was negligible in both the control and CFTR cells. After the addition of forskolin, however, there was a marked rapid cell swelling, after a lag period of from 30 to 90 s, accompanied by elevated intracellular Cl^- concentration in the cells expressing CFTR but not in control cells (Fig. 16E). These results demonstrated that the basis of cAMP-induced Cl^- permeability observed in CFTR-expressing cells was a Cl^- channel.

cAMP-Stimulated Cl^- Currents in the CFTR-Expressing Cells

Corroborating the results of the fluorescence assay, whole-cell current was stimulated in cells expressing CFTR (Fig. 17A). After a lag period of ≈ 30 s, outward current in 11 of 11 cells increased dramatically to a plateau, which was sustained for 2-5 min before decreasing toward control values. This "run down" contrasts with findings using the SPQ assay, possibly reflecting a depletion of cytosolic factors necessary for sustained activation in the whole-cell patch-clamp configuration. currents evoked by the activation cocktail did not display any time-dependent voltage effects (Fig. 17B) contrast to the expressing cells, none of the 9 cells containing the frame-shift CFTR construct

exhibited a response to the activation cocktail (Fig. 17A).

Current-voltage relations were essentially linear in both cell transfected with the intact CFTR and the frame-shift CFTR constructs (now shown). Slope conductances of 1.8 ± 0.4 nS and 1.9 ± 0.5 nS (n = 9) were calculated in control cells before and after treatment with the activation mixture, respectively. The slope conductance in CFTR-expressing cells was similar (1.6 ± 0.1 nS) but treatment with the activation mixture induced an ≈ 13-fold increase to 20.1 ± 1.7 nS (n = 10) (Fig. 17C).

The reversal potential of the cAMP-activated current in cells expressing CFTR was -17 ± 5 mV, approaching the equilibrium potential for Cl^- under standard conditions (E_{Cl} = -32 mV). From this measurement, the calculated anion versus cation permeability was ≈ 5:1. Anion selectivity was assessed in cAMP-activated cells by replacing bath NaCl (135 mM) with sodium gluconate (135 mM). This manipulation resulted in a shift in reversal potential to +33 ± 8 mV (n = 4), toward the predicted E_{Cl} of +41 mV (Fig. 17C).

The inability to construct a full-length CFTR cDNA has hampered progress in understanding the structure and function of the protein. The difficulty is mainly due to the instability of the full-length sequence in E. coli, where there was sequence rearrangement often associated with a short (13 bp) direct repeat in exon 6b. Through the modification of DNA sequence within the first copy of the repeat, however, we have succeeded in the construction of three different plasmids each containing the entire coding sequence. These plasmids replicate efficiently in E. coli (DH5α), and are intact after prolonged propagation. Although this particular modification involves the first repeat, it is understood that this second repeat can also be similarly modified instead of, or in addition to, the first repeat to

achieve similar success in the propagation of the CFTR cDNA.

In addition to verification by direct DNA sequencing, the plasmids have also been examined for their ability to product proteins of expected sized in vitro and in vivo. Based on the longest open reading frame of the consensus cDNA sequence (Riordan et al 1989), a protein of 170 kd (1480 amino acids) has been predicted as the CFTR gene product. The fact that a 170 kd band was detectable in the product translated in vitro and in cells transfected with the full-length CFTR cDNA confirms the original prediction.

The availability of a full-length CFTR cDNA that can be expressed in mammalian and non-mammalian cells offers the opportunity to perform a detailed structure and function analysis of CFTR. The vectors described here are excellent tools for this purpose. With appropriate regulatory sequences inserted upstream of the coding region, it is understood that large quantities of CFTR may be produced through different kinds of heterologous gene expression systems, whereby various biochemical and biophysical studies can be performed.

The ability to express the full-length cDNA also allows development of functional assays for CFTR. In this context, Drumm et al (1990) have demonstrated through a retrovirus-intermediate that the modified full-length cDNA described here was able to confer the function of CFTR in a pancreatic carcinoma cell line (CPFAC-1) derived from a CF patient. Upon proper expression of the cDNA, the cAMP-mediated chloride transport activity was restored in this cell line, providing the first example of functional complementation of CFTR activity. The ability to confer CFTR expression in heterologous cells is an important step towards the possibility of gene therapy in the lung and pancreas of CF patients.

In order to understand the role of individual amino acid residues as well as regions of the protein, site-directed mutagenesis may be used to introduce additional mutations into the coding region of CFTR. The functional assay may also be used to confirm if the sequence alterations detected in CF patients are bona fide disease-causing mutations. The latter consideration is important if broad scale disease diagnosis and carrier screening based on DNA information are to be implemented. Further, since there is a general lack of genotypes, permanent cell lines in which the CF phenotype resists, the ability to generate heterologous cell lines capable of expressing various defective CFTR offers an alternative approach in understanding the function of CFTR and in development of rational therapy. In the latter regard, procedures can be devised for screening of compounds that would interact with the defective protein and restore its function.

To introduce additional mutations into the coding region of CFTR, it is possible to replace regions of the cDNA with altered sequences, as demonstrated by the examples described above. The procedure is difficult, however, because many of the restriction enzyme sites involved are present in more than one vector. For this reason, it is desirable to include unique restriction sites in the coding region of the cDNA. For example, by introducing a silent change (T to C) at position 1944 in exon 13, a novel BstEII site is created at the end of the sequence corresponding to the first NBF1 is probably the most interesting region in CFTR because about one-third of the disease-causing mutations reside in the region. In combination with another unique site, such as SphI, it becomes extremely easy to replace sequences for NBF1.

Finally, the full-length cDNA construct contained in the pBQ4.7 and pBW6.2 may be excised in its entirety by a single PstI digestion or a double digestion with a combination of SaII, XhoI, SmaI or EcoRI. This

versatility allows the cDNA to be transferred from the current vector to other host-vector systems.

References

Cutting, G.R., Kasch, L.M., Rosenstein, B.J., Zielenski, J., Tsui, L.-C., Antonarakis., S.E. and Kazazian, H.H. Jr. (1990) "A Cluster of Cystic Fibrosis Mutations in the First Nucleotide-Binding Fold of the Cystic Fibrosis Conductance Regulator Protein" Nature 346:366-369.

Drumm, M.L., Pope, H.A., Cliff, W.H., Rommens, J.M., Marvin, S.A., Tsui, L.-C., Collins, F.S., Frizzel, R.A. and Wilson, J.M. (1990) "Correction of the Cystic Fibrosis Defect in vitro by Retrovirus-Mediated Gene Transfer" Cell 62:1227-1233

Hawley, D.K. and McClure, W.R., (1983) Nucleic Acids Res. 11:2237-2255

Higuchi, R., Krummel, B., and Saiki, R.K., (1988) "A General Method of in vitro Preparation and Specific Mutagenesis of DNA Fragments: Study of Protein and DNA Interactions", Nucleic Acids Res. 16:7351-7363.

Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K., and Pease, L.R. (1989) "Site-Directed Mutagenesis by Overlap Extension Using the Polymerase Chain Reaction", Gene 77:51-59.

Kerem, B., Rommens, J.M., Buchanan, J.A., Markiewicz, D., Cox, T.K., Chakravarti, A., Buchwald, M. and Tsui, L.-C., (1989) "Identification of the Cystic Fibrosis Gene: Genetic Analysis" Science 245:1073-1080.

Laemmli, U.K., (1970) Nature 227:680-685.

Meakin, S.O., Du, R.P., Tsui, L.-C. and Breitman, M.L. (1987) "γ-Crystallins of the Human Eye Lens: Expression Analysis of Five Members of the Gene Family", Molec. Cell. Biol. 7:2671-2679.

Riordan, J.R., Rommens, J.M., Kerem, B., Alon, N., Rozmahel, R., Grzelchak, K., Zielenski, J., Lok, S., Plavsic, N., Chou J.-L., Drumm, M.L., Iannuzzi, M.C., Collins, F.S., and Tsui, L.-C., (1989) "Identification of the Cystic Fibrosis Gene: Cloning and Characterization of Complementary DNA", Science 245:1055-1073.

Rommens, J.M., Iannuzzi, M.C., Kerem, B., Melmer, G., Drumm, M.L., Melmer, G., Dean, M., Rozmahel, R., Cole J.L., Kennedy, D., Hidaka, N., Zsiga, M., Buchwald, M., Riordan, J.R., Tsui, L.-C., and Collins, F.S., (1989). "Identification of the Cystic Fibrosis Gene: Chromosome Walking and Jumping" Science 245:1059-1065

Saiki, R., Sambrook, J., Fritsch, E.F. and Maniatis, T., (1989) Molecular Cloning Second Edition, Cold Spring Harbor Laboratory Press, NY.

Twobin, H., Staehelin, T. and Gordon J., (1979) Proc. Natl. Acad. Sci USA 76:4350-4354.

Boat, T., Welsh, M.J. & Beaudet, A. (1989) in The Metabolic Basis of Inherited Disease, eds. Scriver, C.R., Beaudet, A.L., Sly. W.S. & Valle, D. (McGraw, Hill, New York), 6th Ed., pp. 2649-2680.

Rommens, J.M., Ianuzzi, M.C., Kerem, B., Drumm, M.L., Melmer, G., Dean, M., Rozmahel, R., Cole, J.L., Kennedy, D., Hidaka, N., Zsiga, M., Burchwald, M., Riordan, J.R., Tsui, L.-C. & Collins. F.S. (1989) Science 245, 1059-1065.

Riordan, J.R., Rommens, J.M., Kerem, B., Alon, N., Rozmahel, R., Grzelczack, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J.-L., Drumm, M.L., Iannuzzi, M.C., Collins, Tsui, L.-C. (1989) Science 245, 1066-1072.

Kerem, B., Riommens, J.M., Buchanan, J.A., Markiewicz, D., Cox, T.K., Chakravarti, A., Buchwald, M. & Tsui, L.-C. (1989) Science 245, 1073-1080).

Dean, M., White, M.B., Amos, J., Gerrard, B, Stewart, C., Khaw, K.-T., & Leppert, M. (1990) Cell 61, 863-870.

Cutting, G.R., Kasch, L.M., Rosenstein, B.J., Zielenski, J., Tsui, L.-C., Kazazian, H.H., Jr., & Antonarakas, S.E. (1990) Nature (London) 346, 366-369.

Kerem, B., Zielenski, J., Markiewicz, D., Bozon, D., Gazit, E., Yahav, J., Kennedy, D., Riordan, J.R., Collins, F.S., Rommens, J.M. & Tsui, L.-C. (1990) Proc. Natl. Acad. Sci. 87, 8447-8451.

White, M.B., Amos, J., Hsu, J.M.C., Gerrard, B., Finn, P. & Dean, M. (1990) Nature (London) 344, 665-667.

Zielenski, J., Bozon, D., Kerem, B., Markiewicz, D., Durie, P., Rommens, J.M. & Tsui, L.-C. (1991) Genomics 10, 229-235.

Hyde, S.C., Emsley, P., Hartshorn, M.J., Mimmack, M.M., Gileadi, U., Pearce, S.R., Gallagher, M.P., Gill, D.R., Hubbard, R.E. & Higgins, C.F. (1990) Nature (London) 346, 362-365.

Gregory, R.J., Cheng, S.H., Rich, D.P., Marshall, J., Paul, S., Hehir, K., Ostedgaard, L., Klinger, K.W., Welsh, M.J. & Smith, A.E. (1990) Nature (London) 347, 382-386.

Drumm, M.L., Pope, H.A., Clift, W.H., Rommens, J.M., Marvin, S.A., Tsui, L.-C., Collins, F.S., Frizzell, R. & Wilson, J.M. (1990) Cell 62, 1227-1233.

Rich, D.P., Anderson, M.P., Gregory, R.J., Cheng, S.H., Paul, S., Jefferson, D.M., McCann, J.D., Klinger, K.W., Smith, A.E. & Welsh, M.J. (1990) Nature (London) 347, 358-363.

Kartner, N., Jensen, T.J., Naismith, A.L., Sun, S., Ackerley, C.A., Reyes, E.F., Tsui, L.-C., Rommens, J.M., Bear, C.E., & Riordan, J.R. (1991) Cell.

Anderson, M.P., Rich, D.P., Gregrory, R.J., Smith, A.E., Welsh, M.J. (1991) Science 251, 679-682.

Meakin, S.O., Du, R.P., Tsui, L.-C. & Breitman, M.L. (1987) Mol. Cell. Biol. 7, 2671-2679.

Karin, M. & Richards, R.I. (1982) Nature (London) 299, 797-802.

Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Lab., Cold Spring Harbor, NY), 2nd Ed.

Higuchi, R., Krummel, B. & Saiki, R.K. (1988) Nucleic Acids Res. 16, 7351-7363.

Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K. & Pease, L.R. (1989) Gene 77, 51-59.

Miller, S.A., Dykes, D.D. & Polesky, H.F. (1988) Nucleic Acids Res. 16, 1215.

MacDonald, R.J., Swift, G.H., Przbyla, A.E. & Chirgwin, J.M. (1987) Methods Enzymol. 152, 219-227.

Laemmli, U.K. (1970) Nature (London) 227, 680-685.

Towbin, H., Staehelin, T. & Gordon, J. (1979) Proc. Natl. Acad. Sci. USA 76, 4350-4354.

Forskett, J.K. (1988) Am. J. Physiol. 255, C566-C571.

Forskett, J.K. & Melvin, J.E. (1989) Science 244, 1582-1585.

Forskett, J.K. (1990) Am. J. Physiol. 259, C998-C1004.

Hamill, O.P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F.J. (1981) Pfügers Arch. 391, 85-100.

Cheng, S.H., Gregory, R.J., Marshall, J., Paul, S., Souza, D.W., White, G.A., O'Riordan, C.R. & Smith, S.E. (1990) Cell 63, 827-834.

Illsley, N.P. & Verkman, A.S. (1987) Biochemistry 26, 1215-1219.

Cliff, W.H. & Frizell, R.A. (1990) Proc. Natl Acad. Sci. USA 87, 4956-4960.

Tabcharani, J.A., Low, W., Elie, D. & Hanrahan, J.W. (1990) FEBS Lett. 270, 157-164.

Gray, M.A., Pollard, C.E., Harris, A., Coleman, L., Greenwell, J.R. & Argent, B.E. (1990) Am. J. Physiol. 259, C752-C761.

Ringe, D. & Petsko, G.A. (1990) Nature (London) 346, 312-313.

THE EMBODIMENTS OF THE INVENTION IN WHICH AN EXCLUSIVE PROPERTY OR PRIVILEGE IS CLAIMED ARE DEFINED AS FOLLOWS:

1. A modified DNA sequence derived from a gene coding for cystic fibrosis transmembrane conductance regulator (CFTR) protein, said gene having at least 27 exons of which normal cDNA codes for CFTR protein, said normal cDNA including exons 6a and 6b wherein exon 6b includes a 13 bp repeat,

said modified DNA sequence comprising at least one of said 13 bp repeats of exon 6b having one or more normal nucleotides of said 13 bp repeat substituted with an alternate nucleotide which continues to code for a corresponding normal amino acid.

2. A modified cDNA sequence of claim 1 wherein said one or more normal nucleotides are at positions 930 and 933 and said substitutions are C for T at 930 and G for A at 933.

3. A DNA construct for use in a recombinant vector comprising said modified cDNA of claim 1.

4. A vector comprising a DNA construct of claim 3, and a promoter sequence for said DNA construct.

5. A vector of claim 4 wherein said vector comprises a restriction map of Figure 5 and 8.

6. A host cell for producing CFTR protein, said host cell comprising a compatible vector of claim 4 whereby expression of said vector in said host cell produces CFTR protein.

7. A host cell of claim 6 selected from the group consisting of bacterial and mammalian cells.

8. A CFTR protein isolated and purified from culture of a host cell of claim 6.

9. A mammalian cell transfected with a vector of claim 4 to enhance Cl^- conductance through cell wall.

FIG.1. 1/20

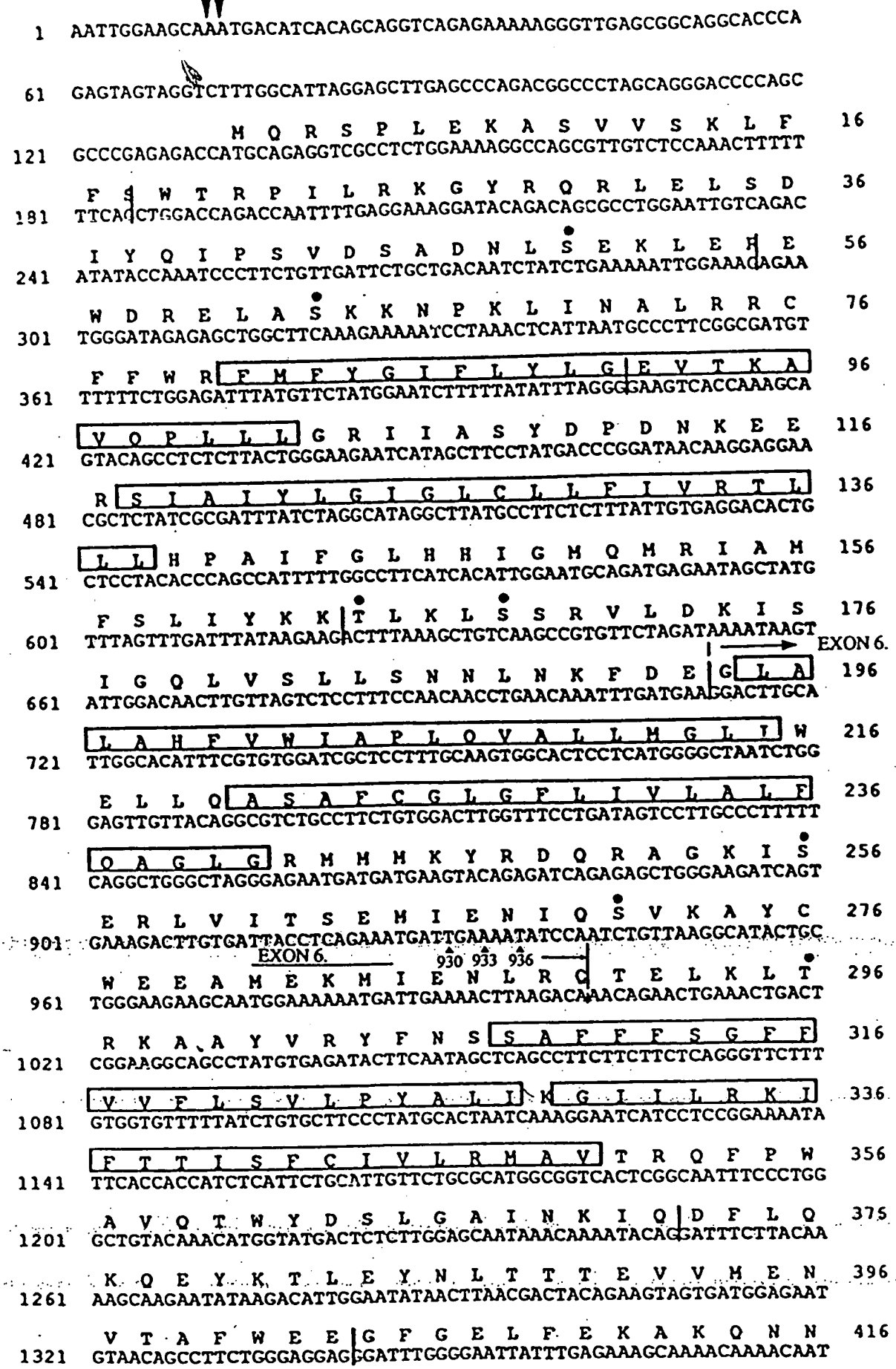


FIG. 1 (cont'd) 2/20

```
          N  N  R  K  T  S  N  G  D  D  S  L  F  F  S  N  F  S  L  L   436
1381  AACAATAGAAAAACTTCTAATGGTGATGACAGCCTCTTCTTCAGTAATTTCTCACTTCTT

          G  T  P  V  L  K  D  I  N  F  K  I  E  R  G  Q  L  L  A  V   456
1441  GGTACTCCTGTCCTGAAAGATATTAATTTCAAGATAGAAAGAGGACAGTTGTTGGCGGTT

          A  G  S  T  G  A  G  K  T  S  L  L  M  M  I  M  G  E  L  E   476
1501  GCTGGATCCACTGGAGCAGGCAAGACTTCACTTCTAATGATGATTATGGGAGAACTGGAG

          P  S  E  G  K  I  K  H  S  G  R  I  S  F  C  S  Q  F  S  W   496
1561  CCTTCAGAGGGTAAAATTAAGCACAGTGGAAGAATTTCATTCTGTTCTCAGTTTTCCTGG

          I  M  P  G  T  I  K  E  N  I  I  F  G  V  S  Y  D  E  Y  R   516
1621  ATTATGCCTGGCACCATTAAAGAAAATATCATCTTTGGTGTTTCCTATGATGAATATAGA

          Y  R  S  V  I  K  A  C  Q  L  E  E  D  I  S  K  F  A  E  K   536
1681  TACAGAAGCGTCATCAAAGCATGCCAACTAGAAGAGGACATCTCCAAGTTTGCAGAGAAA

          D  N  I  V  L  G  E  G  G  I  T  L  S  G  G  Q  R  A  R  I   556
1741  GACAATATAGTTCTTGGAGAAGGTGGAATCACACTGAGTGGAGGTCAACGAGCAAGAATT

          S  L  A  R  A  V  Y  K  D  A  D  L  Y  L  L  D  S  P  F  G   576
1801  TCTTTAGCAAGAGCAGTATACAAAGATGCTGATTTGTATTTATTAGACTCTCCTTTTGGA

          Y  L  D  V  L  T  E  K  E  I  F  E  S  C  V  C  K  L  M  A   596
1861  TACCTAGATGTTTTAACAGAAAAAGAAATATTTGAAAGCTGTGTCTGTAAACTGATGGCT

          N  K  T  R  I  L  V  T  S  K  M  E  H  L  K  K  A  D  K  I   616
1921  AACAAAACTAGGATTTTGGTCACTTCTAAAATGGAACATTTAAAGAAAGCTGACAAAATA

          L  I  L  H  E  G  S  S  Y  F  Y  G  T  F  S  E  L  Q  N  L   636
1981  TTAATTTTGCATGAAGGTAGCAGCTATTTTTATGGGACATTTTCAGAACTCCAAAATCTA

          Q  P  D  F  S  S  K  L  M  G  C  D  S  F  D  Q  F  S  A  E   656
2041  CAGCCAGACTTTAGCTCAAAACTCATGGGATGTGATTCTTTCGACCAATTTAGTGCAGAA

          R  R  N  S  I  L  T  E  T  L  H  R  F  S  L  E  G  D  A  P   676
2101  AGAAGAAATTCAATCCTAACTGAGACCTTACACCGTTTCTCATTAGAAGGAGATGCTCCT

          V  S  W  T  E  T  K  K  Q  S  F  K  Q  T  G  E  F  G  E  K   696
2161  GTCTCCTGGACAGAAACAAAAAAACAATCTTTTAAACAGACTGGAGAGTTTGGGGAAAAA

          R  K  N  S  I  L  N  P  I  N  S  I  R  K  F  S  I  V  Q  K   716
2221  AGGAAGAATTCTATTCTCAATCCAATCAACTCTATACGAAAATTTTCCATTGTGCAAAAG

          T  P  L  Q  M  N  G  I  E  E  D  S  D  E  P  L  E  R  R  L   736
2281  ACTCCCTTACAAATGAATGGCATCGAAGAGGATTCTGATGAGCCTTTAGAGAGAAGGCTG

          S  L  V  P  D  S  E  Q  G  E  A  I  L  P  R  I  S  V  I  S   756
2341  TCCTTAGTACCAGATTCTGAGCAGGGAGAGGCGATACTGCCTCGCATCAGCGTGATCAGC

          T  G  P  T  L  Q  A  R  R  R  Q  S  V  L  N  L  M  T  H  S   776
2401  ACTGGCCCCACGCTTCAGGCACGAAGGAGGCAGTCTGTCCTGAACCTGATGACACACTCA

          V  N  Q  G  Q  N  I  H  R  K  T  T  A  S  T  R  K  V  S  L   796
2461  GTTAACCAAGGTCAGAACATTCACCGAAAGACAACAGCATCCACACGAAAAGTGTCACTG

          A  P  Q  A  N  L  T  E  L  D  I  Y  S  R  R  L  S  Q  E  T   816
2521  GCCCCTCAGGCAAACTTGACTGAACTGGATATATATTCAAGAAGGTTATCTCAAGAAACT
```

FIG.1 (cont'd) 3/20

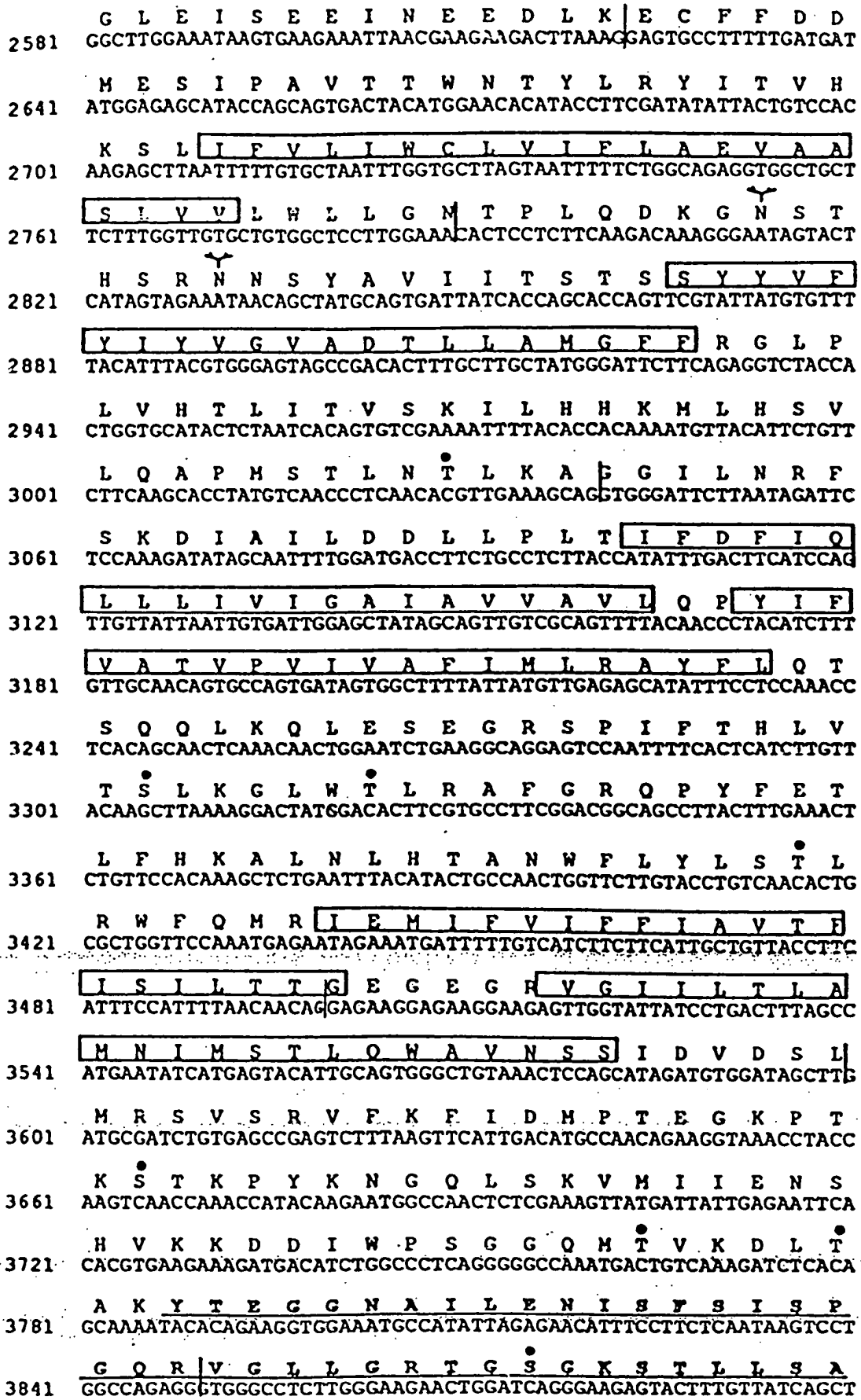


FIG.1.(cont'd) 4/20

```
          F  L  R  L  L  N  T  E  G  E  I  Q  I  D  G  V  S  W  D  S   1276
3901  TTTTTGAGACTACTGAACACTGAAGGAGAAATCCAGATCGATGGTGTGTCTTGGGATTCA

          I  T  L  Q  Q  W  R  K  A  F  G  V  I  P  Q  K  V  F  I  F   1296
3961  ATAACTTTGCAACAGTGGAGGAAAGCCTTTGGAGTGATACCACAGAAAGTATTTATTTTT

          S  G  T  F  R  K  N  L  D  P  Y  E  Q  W  S  D  Q  E  I  W   1316
4021  TCTGGAACATTTAGAAAAAACTTGGATCCCTATGAACAGTGGAGTGATCAAGAAATATGG

          K  V  A  D  E  V  G  L  R  S  V  I  E  Q  F  P  G  K  L  D   1336
4081  AAAGTTGCAGATGAGGTTGGGCTCAGATCTGTGATAGAACAGTTTCCTGGGAAGCTTGAC

          F  V  L  V  D  G  G  C  V  L  S  H  G  H  K  Q  L  M  C  L   1356
4141  TTTGTCCTTGTGGATGGGGGCTGTGTCCTAAGCCATGGCCACAAGCAGTTGATGTGCTTG

          A  R  S  V  L  S  K  A  K  I  L  L  L  D  E  P  S  A  H  L   1376
4201  GCTAGATCTGTTCTCAGTAAGGCGAAGATCTTGCTGCTTGATGAACCCAGTGCTCATTTG

          D  P  V  T  Y  Q  I  I  R  R  T  L  K  Q  A  F  A  D  C  T   1396
4261  GATCCAGTAACATACCAAATAATTAGAAGAACTCTAAAACAAGCATTTGCTGATTGCACA

          V  I  L  C  E  H  R  I  E  A  M  L  E  C  Q  Q  F  L  V  I   1416
4321  GTAATTCTCTGTGAACACAGGATAGAAGCAATGCTGGAATGCCAACAATTTTTGGTCATA

          E  E  N  K  V  R  Q  Y  D  S  I  Q  K  L  L  N  E  R  S  L   1436
4381  GAAGAGAACAAAGTGCGGCAGTACGATTCCATCCAGAAACTGCTGAACGAGAGGAGCCTC

          F  R  Q  A  I  S  P  S  D  R  V  K  L  F  P  H  R  N  S  S   1456
4441  TTCCGGCAAGCCATCAGCCCCTCCGACAGGGTGAAGCTCTTTCCCCACCGGAACTCAAGC

          K  C  K  S  K  P  Q  I  A  A  L  K  E  E  T  E  E  E  V  Q   1476
4501  AAGTGCAAGTCTAAGCCCCAGATTGCTGCTCTGAAAGAGGAGACAGAAGAAGAGGTGCAA

          D  T  R  L  -                                               1480
4561  GATACAAGGCTTTAGAGAGCAGCATAAATGTTGACATGGGACATTTGCTCATGGAATTGG
4621  AGCTCGTGGGACAGTCACCTCATGGAATTGGAGCTCGTGGAACAGTTACCTCTGCCTCAG
4681  AAAACAAGGATGAATTAAGTTTTTTTTTAAAAAAGAAACATTTGGTAAGGGGAATTGAGG
4741  ACACTGATATGGGTCTTGATAAATGGCTTCCTGGCAATAGTCAAATTGTGTGAAAGGTAC
4801  TTCAAATCCTTGAAGATTTACCACTTGTGTTTTGCAAGCCAGATTTTCCTGAAAACCCTT
4861  GCCATGTGCTAGTAATTGGAAAGGCAGCTCTAAATGTCAATCAGCCTAGTTGATCAGCTT
4921  ATTGTCTAGTGAAACTCGTTAATTTGTAGTGTTGGAGAAGAACTGAAATCATACTTCTTA
4981  GGGTTATGATTAAGTAATGATAACTGGAAACTTCAGCGGTTTATATAAGCTTGTATTCCT
5041  TTTTCTCTCCTCTCCCCATGATGTTTAGAAACACAACTATATTGTTTGCTAAGCATTCCA
5101  ACTATCTCATTTCCAAGCAAGTATTAGAATACCACAGGAACCACAAGACTGCACATCAAA
5161  ATATGCCCCATTCAACATCTAGTGAGCAGTCAGGAAAGAGAACTTCCAGATCCTGGAAAT
5221  CAGGGTTAGTATTGTCCAGGTCTACCAAAAATCTCAATATTTCAGATAATCACAATACAT
5281  CCCTTACCTGGGAAAGGGCTGTTATAATCTTTCACAGGGGACAGGATGGTTCCCTTGATG
5341  AAGAAGTTGATATGCCTTTTCCCAACTCCAGAAAGTGACAAGCTCACAGACCTTTGAACT
5401  AGAGTTTAGCTGGAAAAGTATGTTAGTGCAAATTGTCACAGGACAGCCCTTCTTTCCACA
5461  GAAGCTCCAGGTAGAGGGTGTGTAAGTAGATAGGCCATGGGCACTGTGGGTAGACACACA
5521  TGAAGTCCAAGCATTTAGATGTATAGGTTGATGGTGGTATGTTTTCAGGCTAGATGTATG
5581  TACTTCATGCTGTCTACACTAAGAGAGAATGAGAGACACACTGAAGAAGCACCAATCATG
5641  AATTAGTTTTATATGCTTCTGTTTTATAATTTTGTGAAGCAAAATTTTTTCTCTAGGAAA
5701  TATTTATTTTAATAATGTTTCAAACATATATTACAATGCTGTATTTTAAAAGAATGATTA
5761  TGAATTACATTTGTATAAAATAATTTTTATATTTGAAATATTGACTTTTTATGGCACTAG
5821  TATTTTTATGAAATATTATGTTAAAACTGGGACAGGGGAGAACCTAGGGTGATATTAACC
5881  AGGGGCCATGAATCACCTTTTGGTCTGGAGGGAAGCCTTGGGGCTGATCGAGTTGTTGCC
5941  CACAGCTGTATGATTCCCAGCCAGACACAGCCTCTTAGATGCAGTTCTGAAGAAGATGGT
6001  ACCACCAGTCTGACTGTTTCCATCAAGGGTACACTGCCTTCTCAACTCCAAACTGACTCT
6061  TAAGAAGACTGCATTATATTTATTACTGTAAGAAAATATCACTTGTCAATAAAATCCATA
6121  CATTTGTGT(A)n
```

FIG. 2a. CF GENE

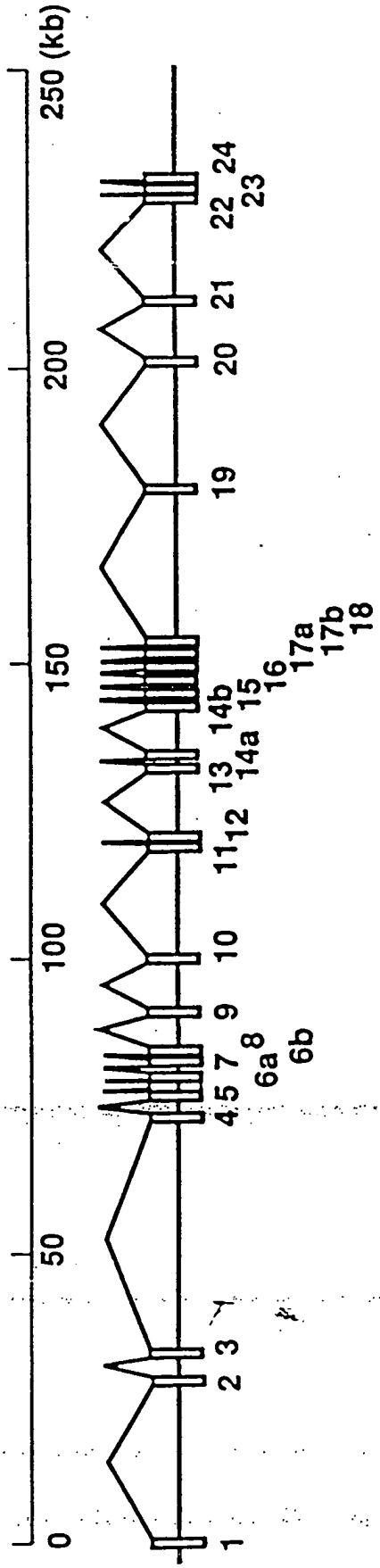


FIG. 2b. CFTR

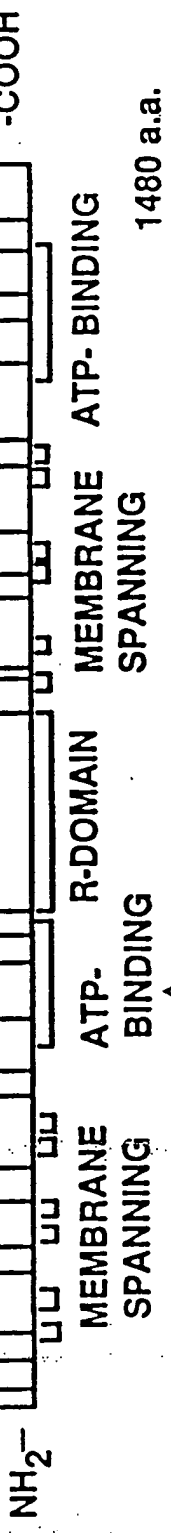


FIG.3.

```
                876
Exon 6b:        AGATCAGAGAGCTGGGAAGATCAGTGAAAGACTTGTGATTACCTCAG

                -35                                  -10
                TTGACA----------(17±1 bp)----------TATAAT
AAATGATTGAAAAATATCCAATCTGTTAAGGCATACTGCTGGGAAGAAGCAATGGAAA
                  C   G   C
                930 933 936

                                                        1027
AAATGATTGAAAACTTAAGACAAACAGAACTGAACTGACTCGGAAGG
```

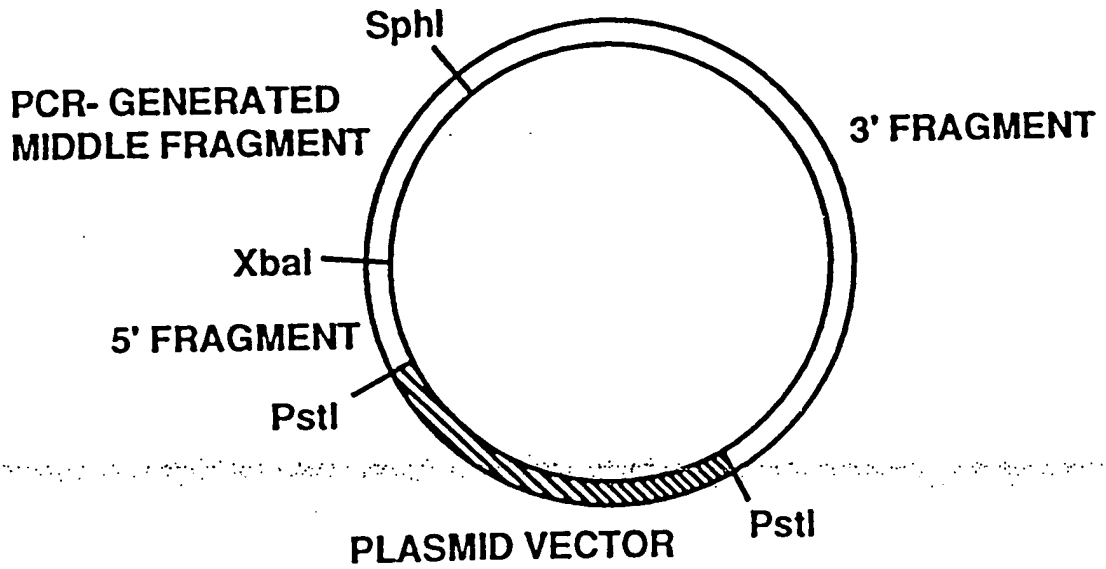

SphI
PCR- GENERATED
MIDDLE FRAGMENT
3' FRAGMENT
XbaI
5' FRAGMENT
PstI
PstI
PLASMID VECTOR

FIG. 4.

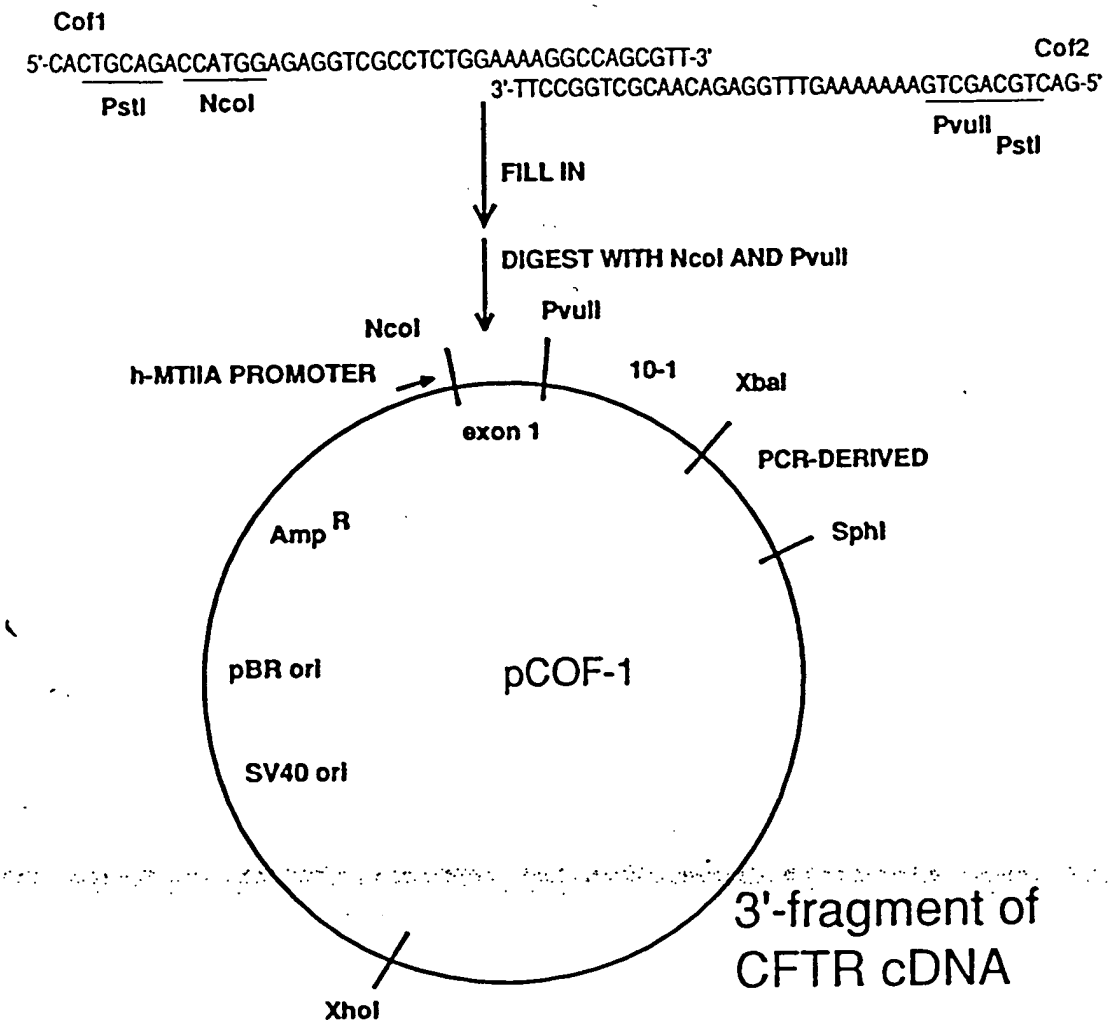

Cof1
5'-CACTGCAGACCATGGAGAGGTCGCCTCTGGAAAAGGCCAGCGTT-3'
PstI
NcoI
Cof2
3'-TTCCGGTCGCAACAGAGGTTTGAAAAAAAGTCGACGTCAG-5'
PvuII
PstI
FILL IN
DIGEST WITH NcoI AND PvuII
NcoI
PvuII
h-MTIIA PROMOTER
10-1
XbaI
exon 1
PCR-DERIVED
Amp R
SphI
pBR ori
pCOF-1
SV40 ori
XhoI
3'-fragment of
CFTR cDNA

FIG. 5.

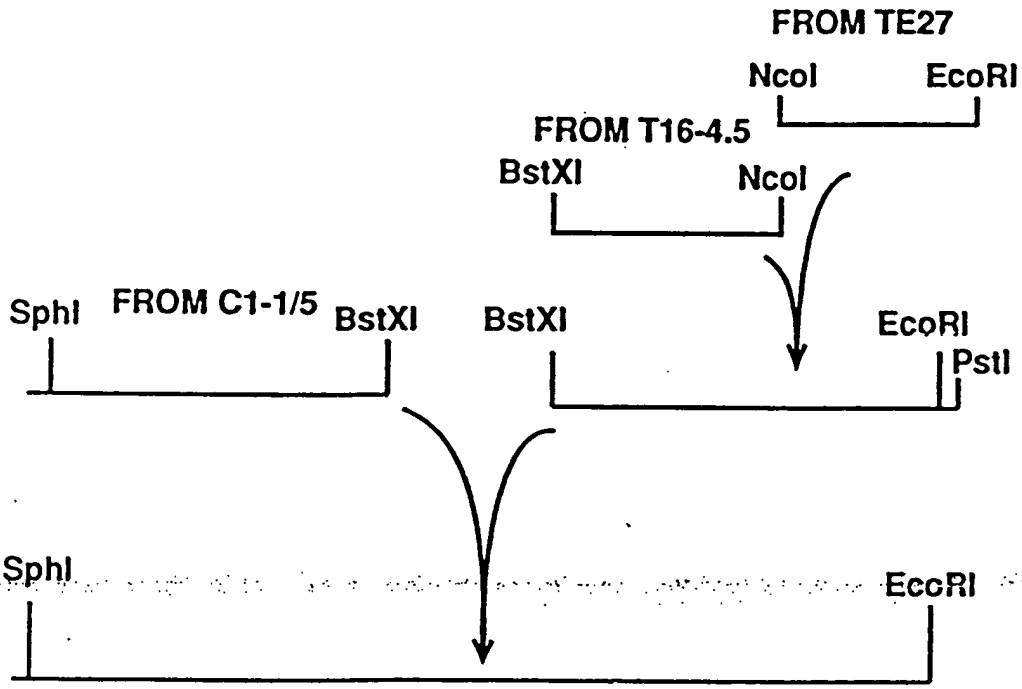

FROM TE27
NcoI
EcoRI
FROM T16-4.5
BstXI
NcoI
SphI
FROM C1-1/5
BstXI
BstXI
EcoRI
PstI
SphI
EcoRI

FIG. 6.

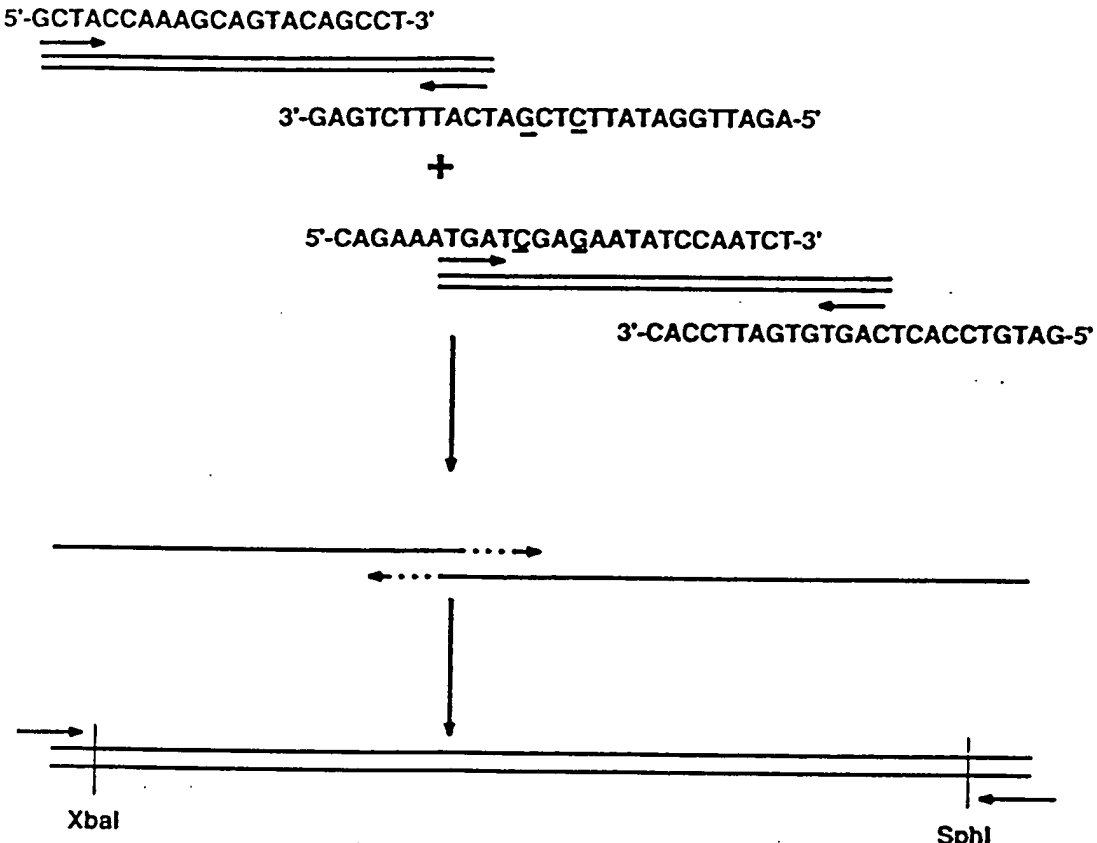

5'-GCTACCAAAGCAGTACAGCCT-3'
3'-GAGTCTTTACTAGCTCTTATAGGTTAGA-5'
+
5'-CAGAAATGATCGAGAATATCCAATCT-3'
3'-CACCTTAGTGTGACTCACCTGTAG-5'
XbaI
SphI

FIG.7.

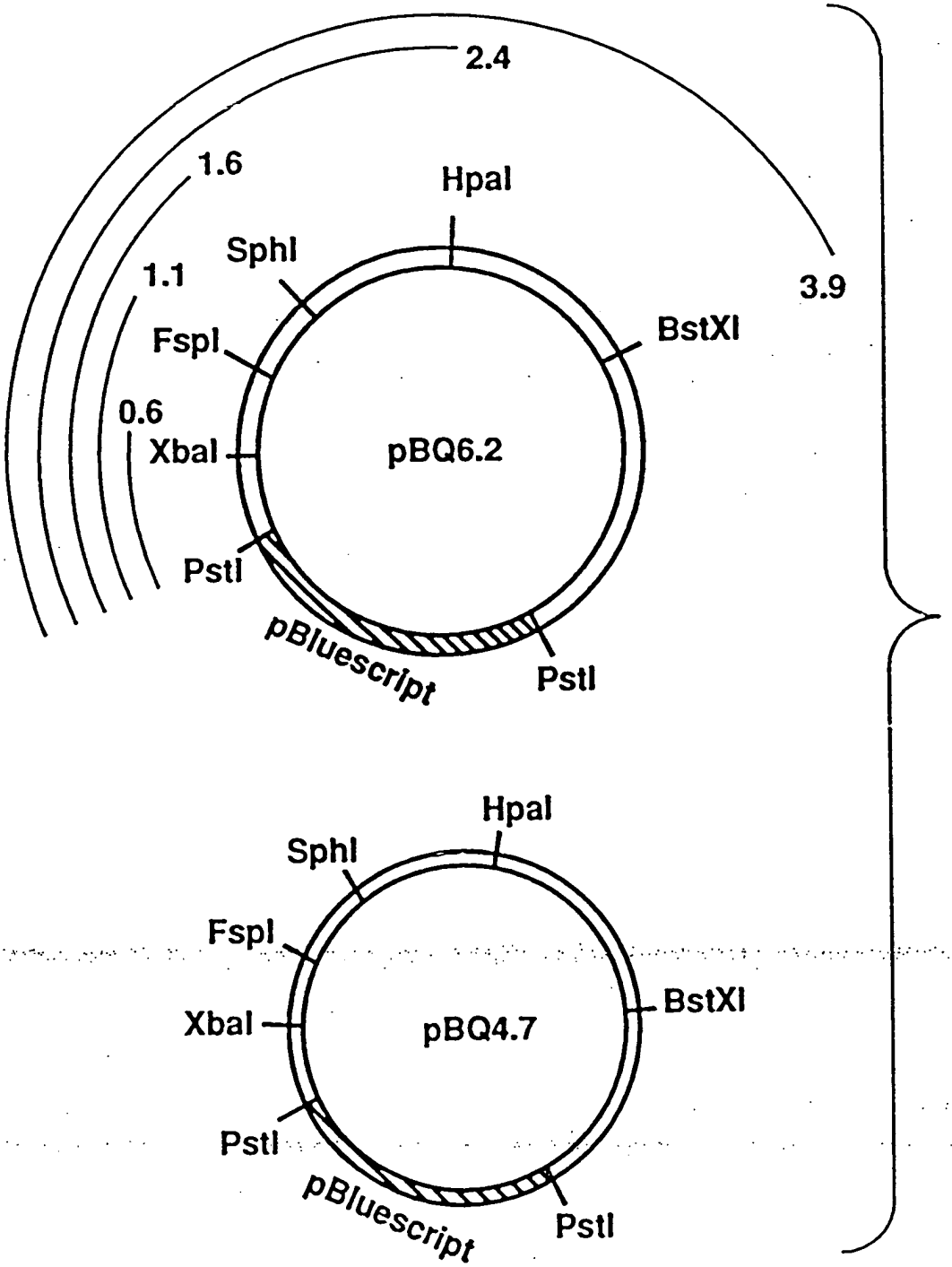

2.4
1.6
1.1
0.6
3.9
HpaI
SphI
FspI
XbaI
PstI
BstXI
pBQ6.2
pBluescript
PstI
HpaI
SphI
FspI
XbaI
PstI
BstXI
pBQ4.7
pBluescript
PstI

FIG. 8.

Sequence of Exon 6b in Constructed Vectors

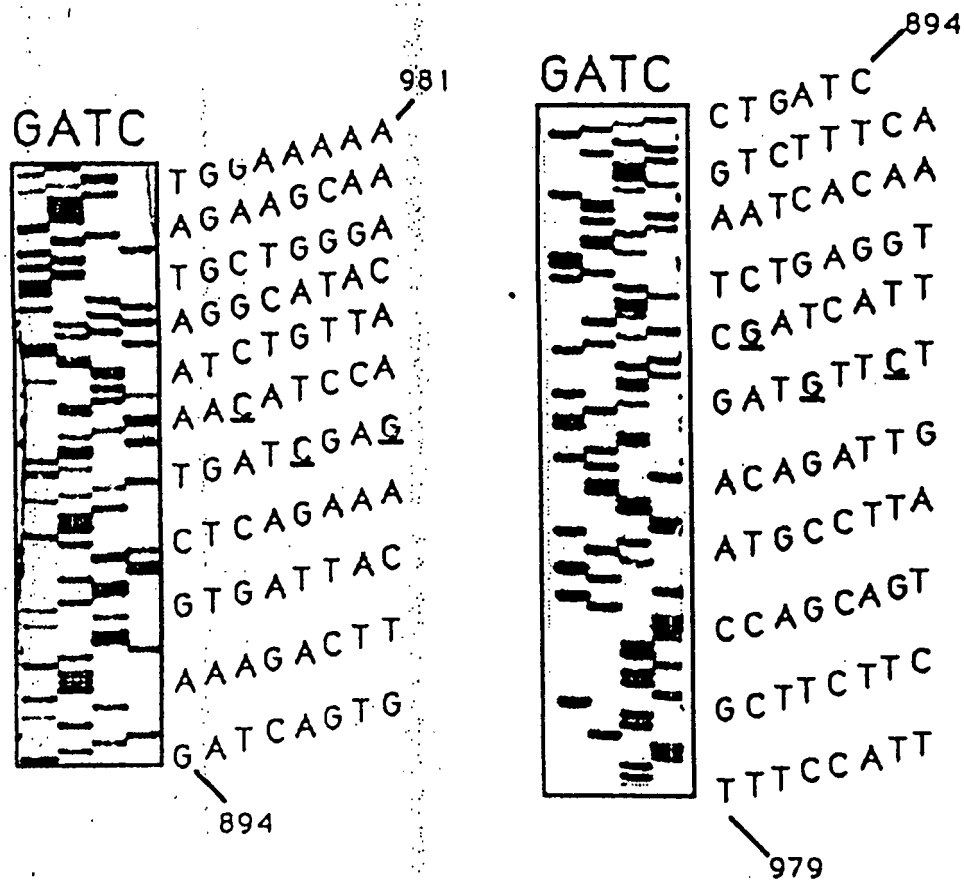


FIG.9.

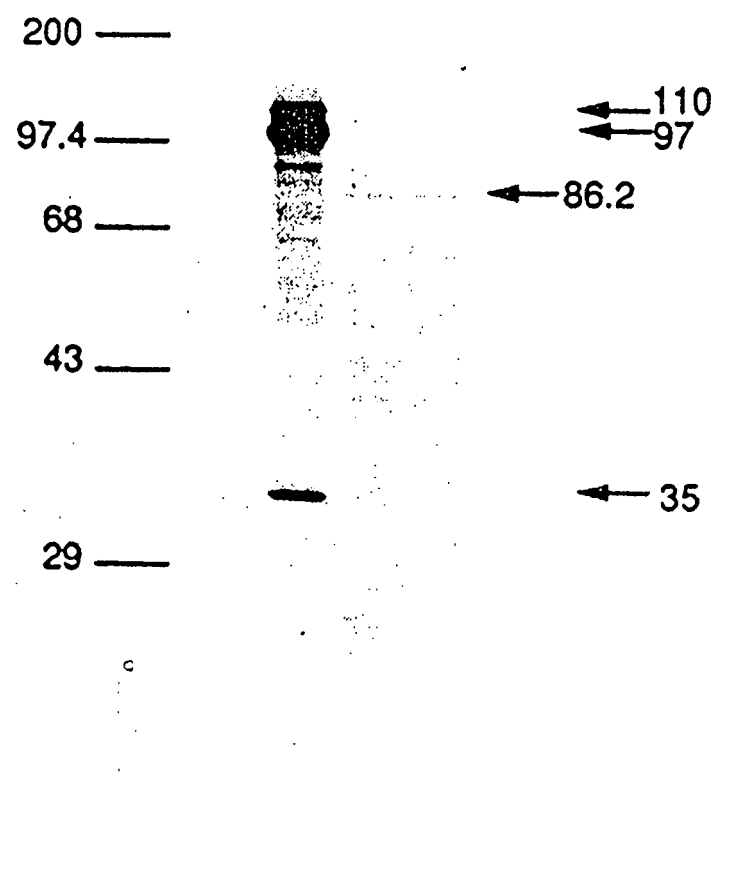


FIG.10.

DNA ANALYSIS

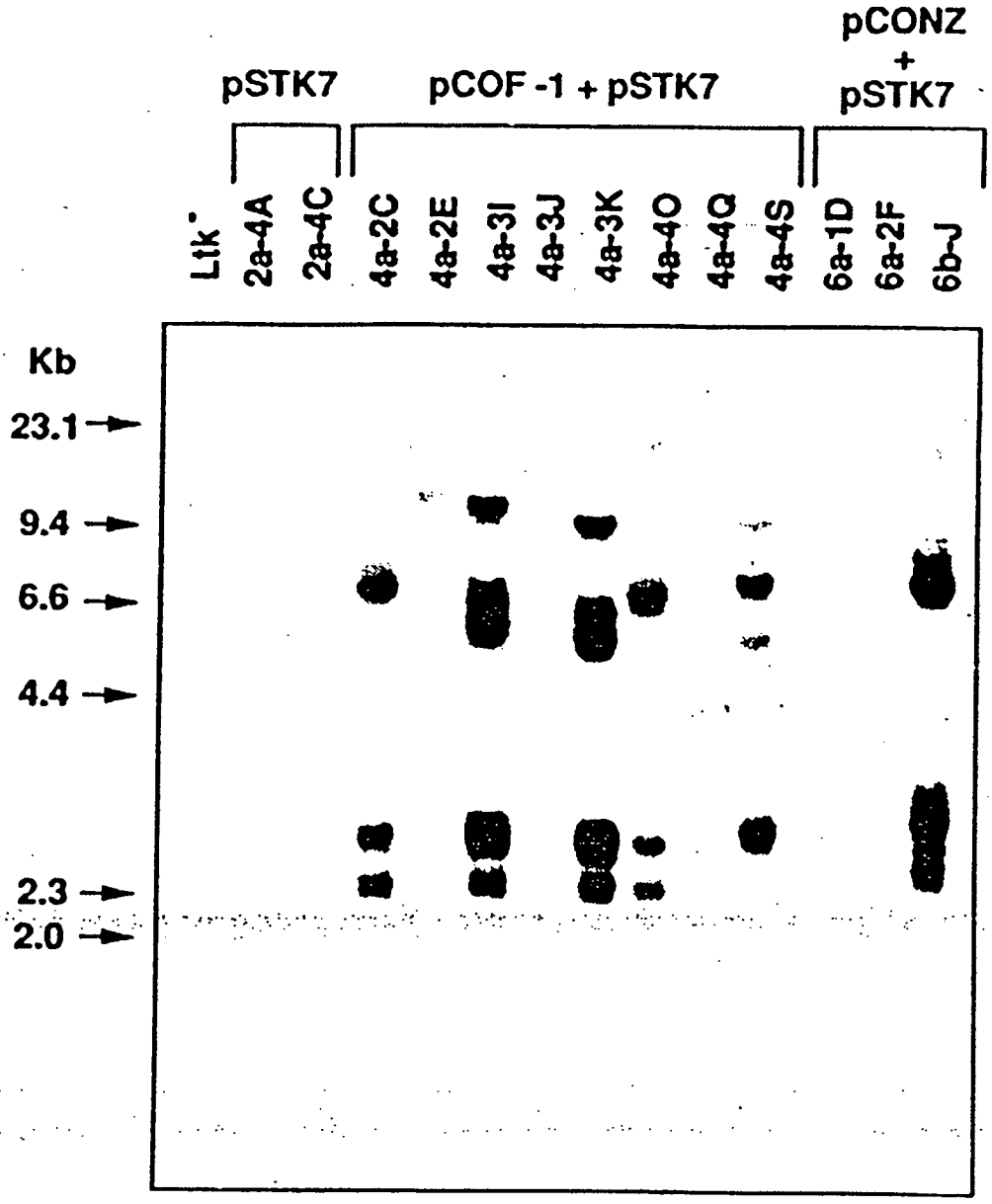
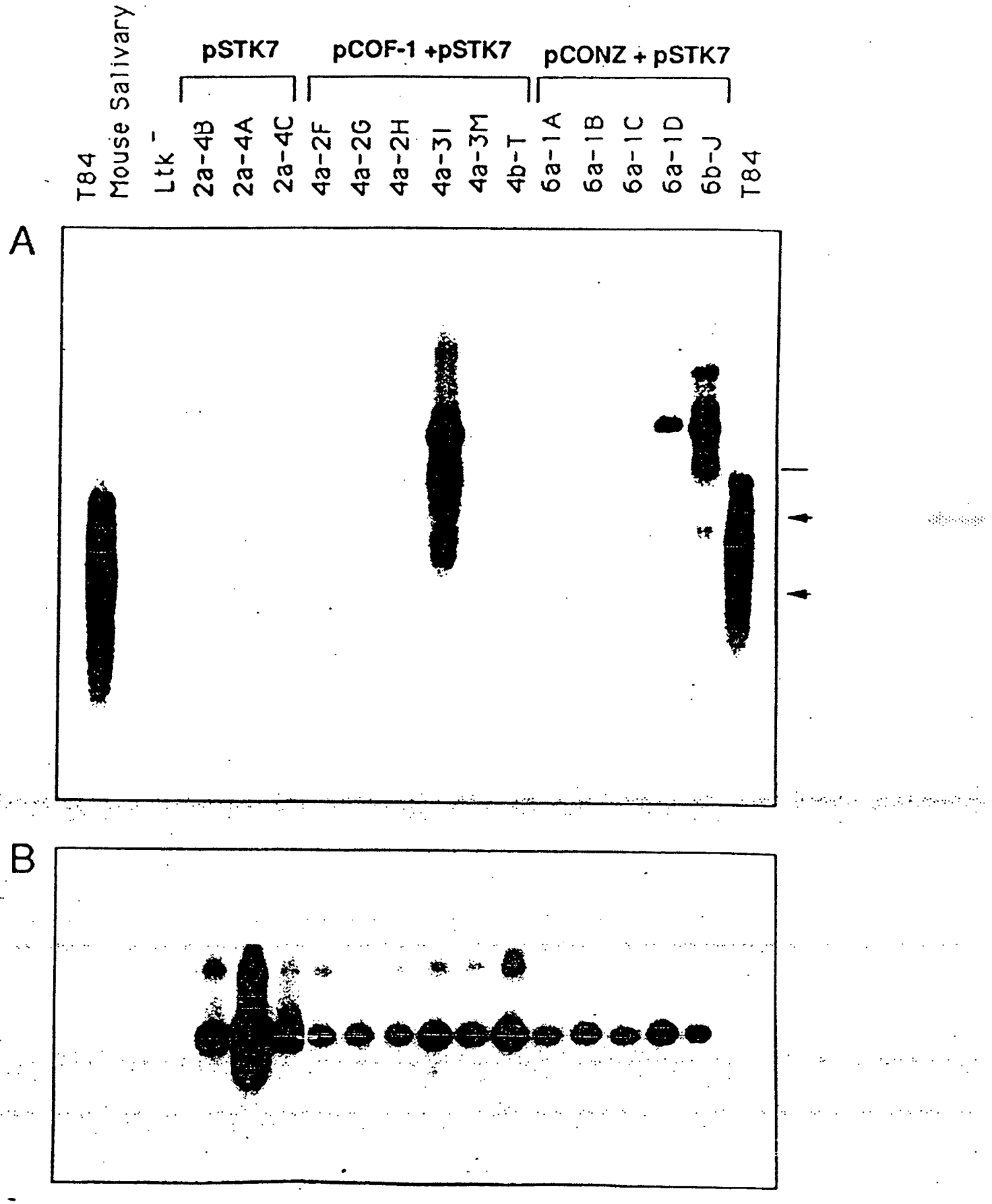


FIG.11.

RNA ANALYSIS

FIG.12



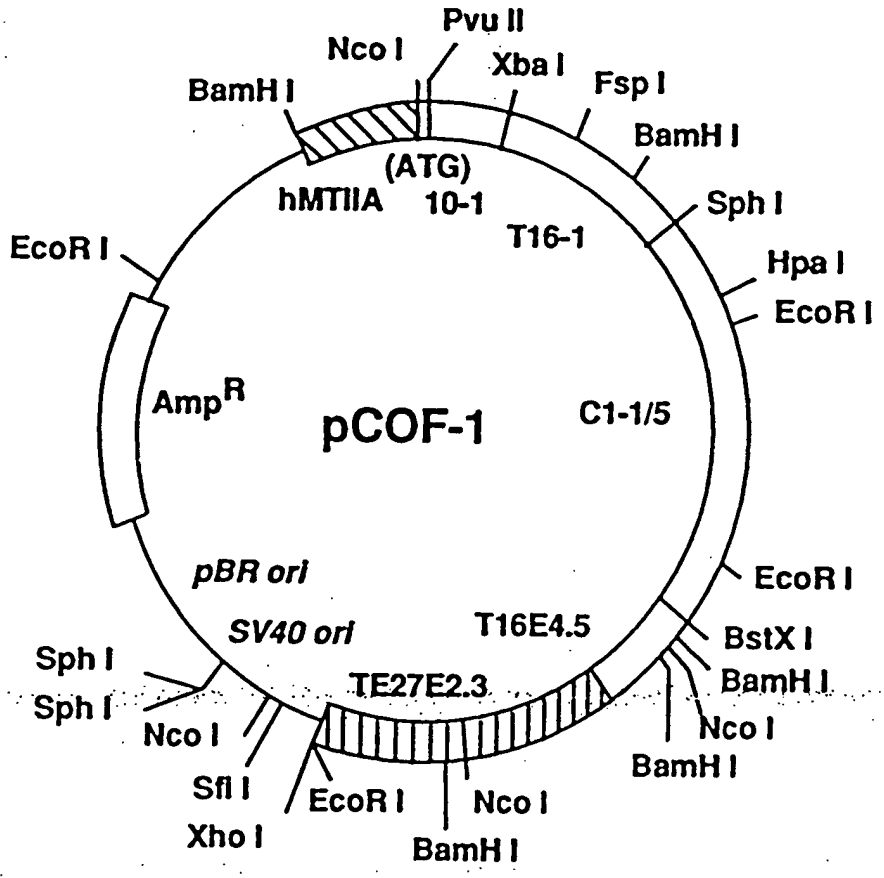

Pvu II
Nco I
Xba I
Fsp I
BamH I
BamH I
(ATG)
hMTIIA
10-1
Sph I
T16-1
Hpa I
EcoR I
EcoR I
Amp^R
pCOF-1
C1-1/5
pBR ori
SV40 ori
T16E4.5
EcoR I
BstX I
BamH I
Nco I
BamH I
TE27E2.3
Sph I
Sph I
Nco I
Sfl I
Xho I
EcoR I
Nco I
BamH I

FIGURE 14.

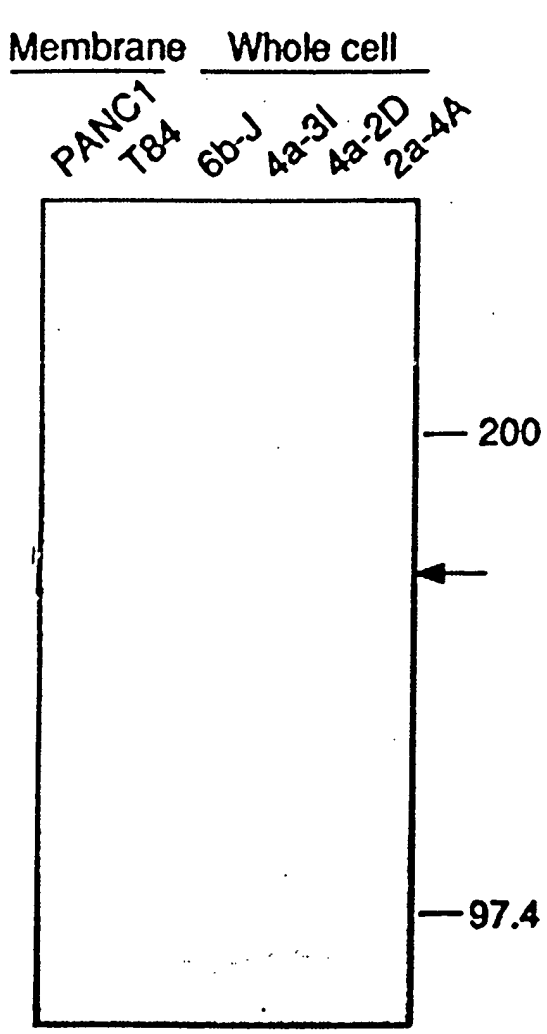
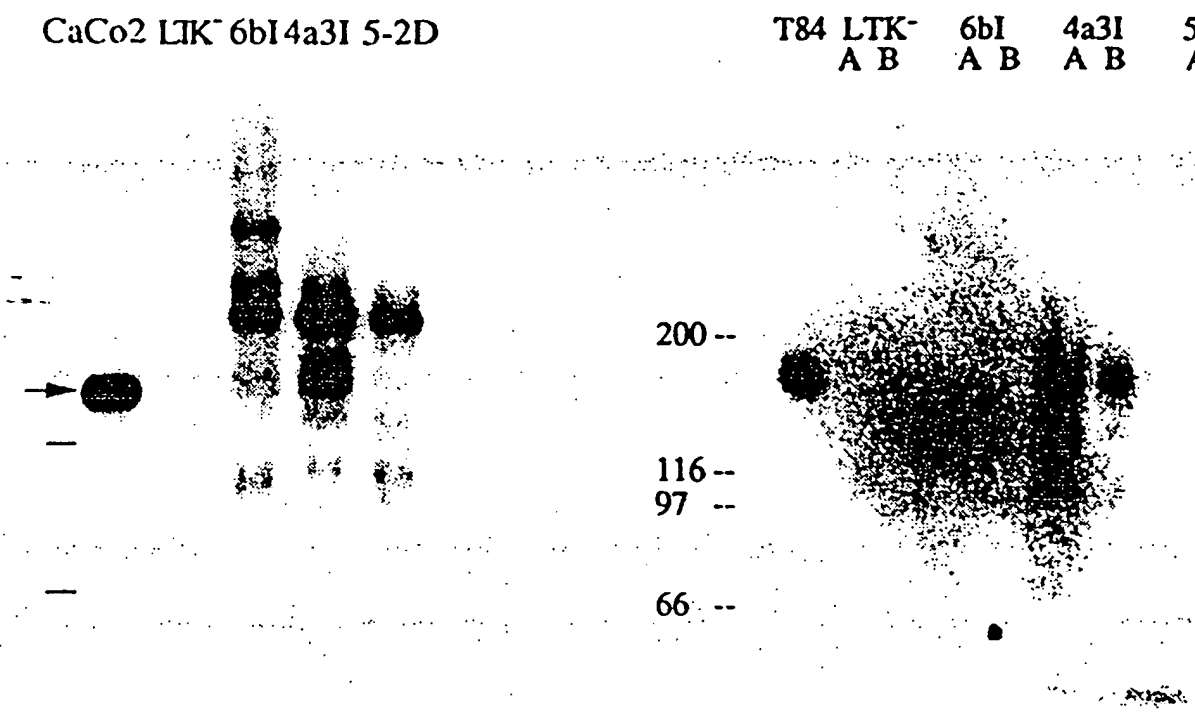

Membrane
Whole cell
PANC1
T84
6b-J
4a-3I
4a-2D
2a-4A
200
97.4

FIG.13.

FIG.15.


CaCo2 LTK- 6bI 4a3I 5-2D
T84 LTK- 6bI 4a3I 5-2D
A B A B A B A B
200 --
116 --
97 --
66 --

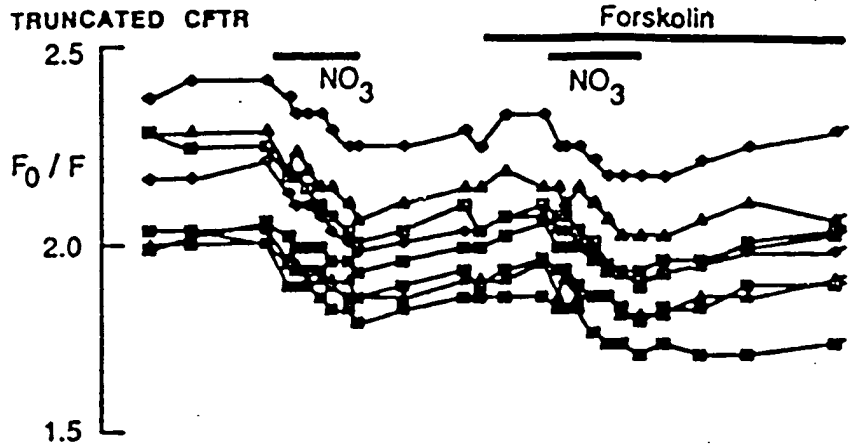


FIG.16A.

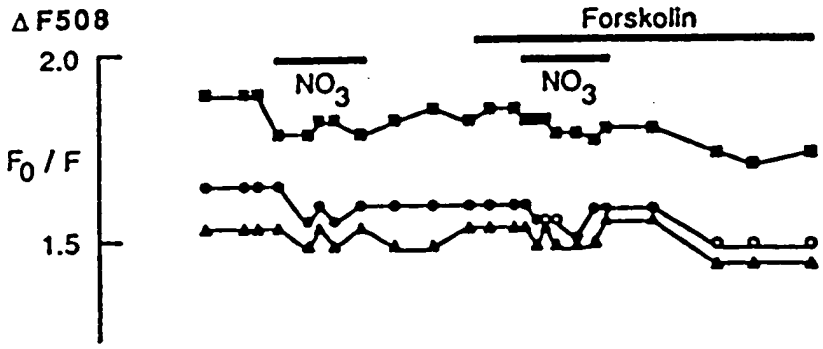


FIG.16B.

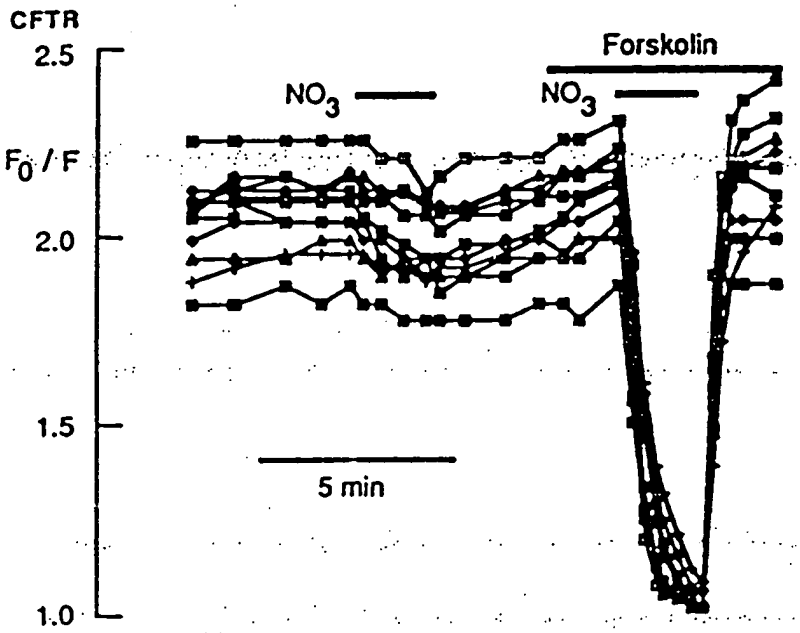


FIG.16C.

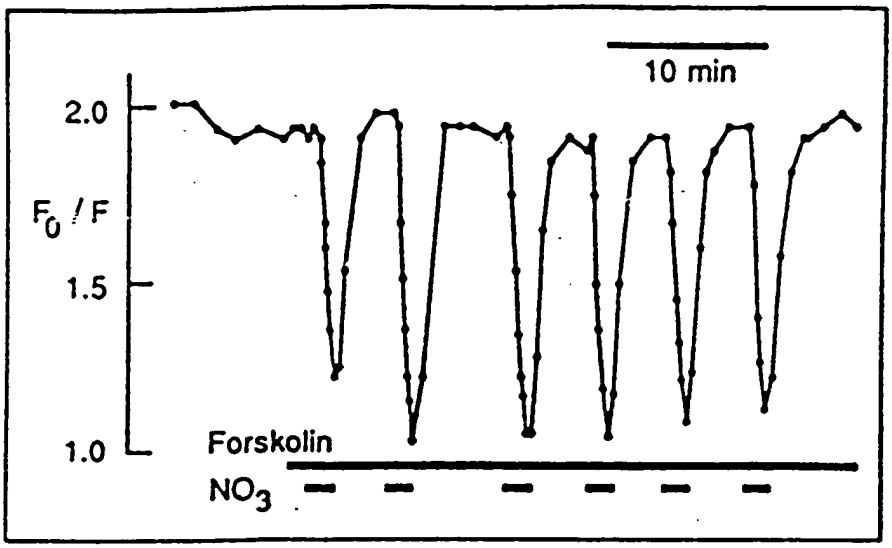


FIG.16D.

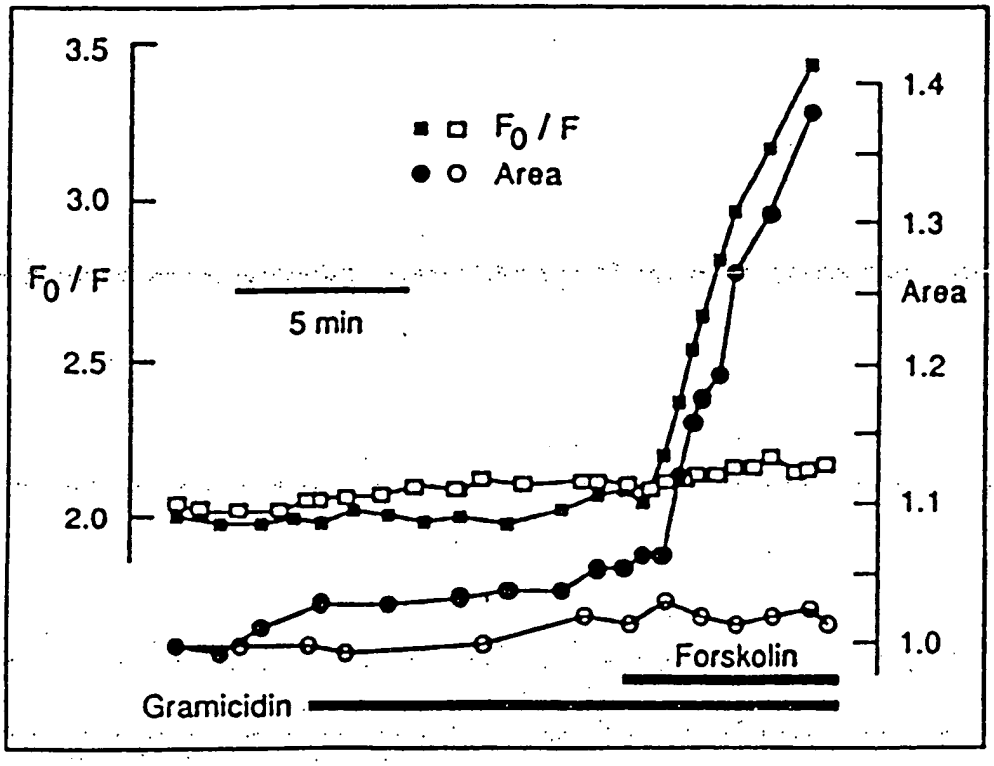


FIG.16E.

FIG. 17A.

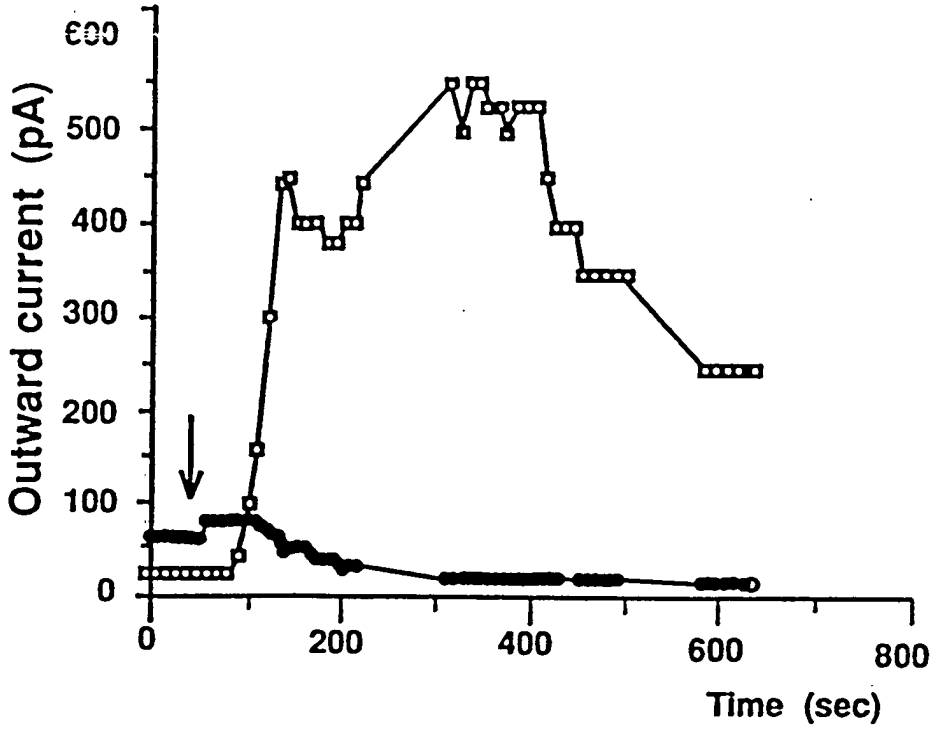
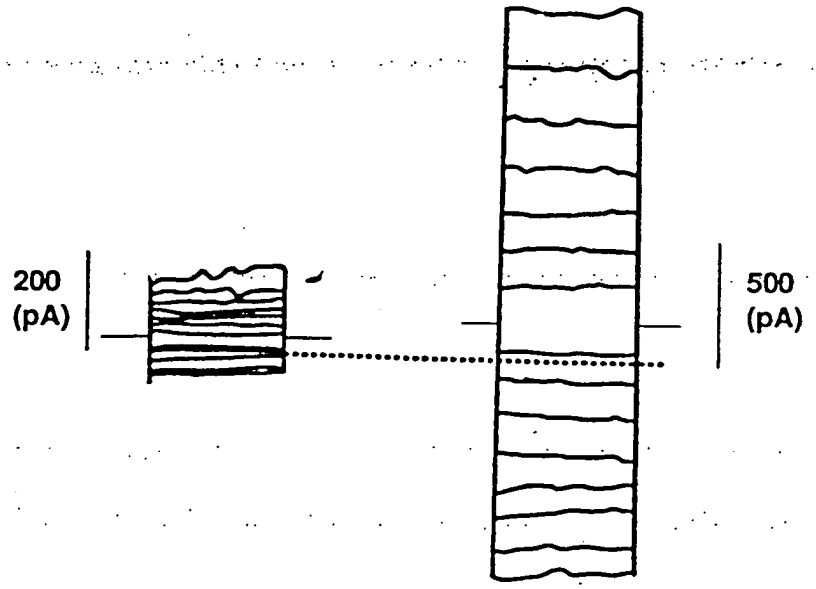


FIG.17B.



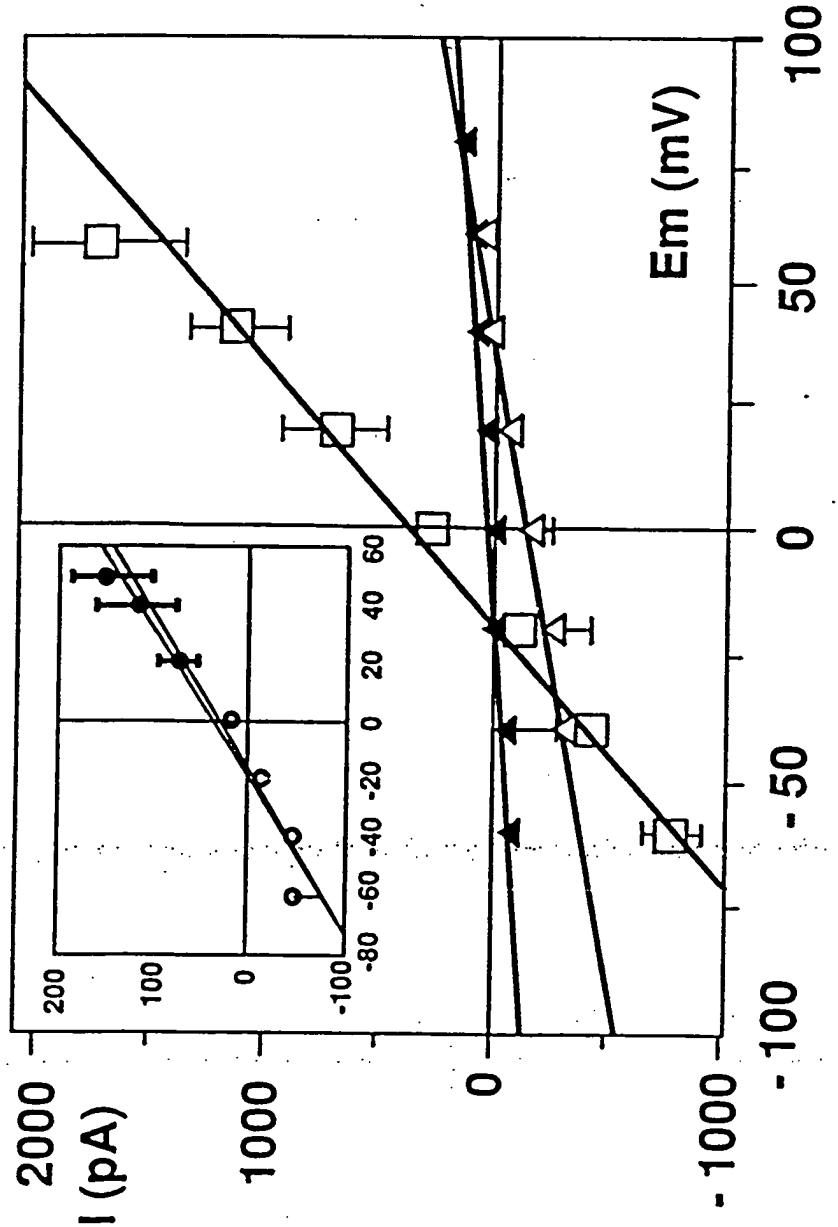

Em (mV)
100
50
0
-50
-100
I (pA)
2000
1000
0
-1000
200
100
0
-100
60
40
20
0
-20
-40
-60
-80

FIG. 17C.

INTERNATIONAL SEARCH REPORT

International Application No PCT/CA 91/00341

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all)[6]

According to International Patent Classification (IPC) or to both National Classification and IPC

Int.Cl.5 C 12 N 15/12 C 12 N 1/21 C 12 N 5/16
C 07 K 13/00

II. FIELDS SEARCHED

Minimum Documentation Searched[7]

Classification System	Classification Symbols
Int.Cl.5	C 12 N C 07 K

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched[8]

III. DOCUMENTS CONSIDERED TO BE RELEVANT[9]

Category °	Citation of Document,[11] with indication, where appropriate, of the relevant passages[12]	Relevant to Claim No.[13]
P,X	Cell, volume 62, no. 6, 21 September 1990, Cell Press, (Cambridge, NA, GB) M.L. Drumm et al.: "Correction of the cystic fibrosis defect in vitro by retrovirus-mediated gene transfer", pages 1227-1233, see page 1228, left-hand column, line 1 - right-hand column, line 9; page 1229, left-hand column, line 22 - page 1231, left-hand column, line 10 (cited in the application) ---	1-4,6,7 ,9
P,Y	---	5
Y	Mol. Cell. Biol., volume 7, no. 8, August 1987, American Society for Microbiology (Washington, US) S.O. Meakin et al.: "Gamma-crystallins of the human eye lens: expression analysis of five members of the gene family", pages 2671-2679, see figure 6 (cited in the application) --- -/-	5

° Special categories of cited documents :[10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
12-11-1991	19. 12. 91
International Searching Authority	Signature of Authorized Officer
EUROPEAN PATENT OFFICE	[signature]

Form PCT/ISA/210 (second sheet) (January 1985)

Mme Dagmar FRANK

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category °	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
Y	Nucleic Acids Research, volume 16, no. 15, 11 August 1988, IRL Press LTD (Oxford, GB) J.M. Short et al.: "Lambda ZAP: a bacteriophage lambda expression vector with in vivo excision properties", pages 7583-7600, pages 7583-7600, see figure 2; page 7585, line 27 - page 7588, line 3	5
A	Science, volume 245, 8 September 1989, AAAS (Washington, DC, US) J.R. Riordan et al.: "Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA", pages 1066-1073, see figures 1,2 (cited in the application)	1-9

8

THE FULL-LENGTH CFTR cDNA IS TOXIC IN BACTERIAL HOSTS. M.Drumm, K. Murphy, H. Pope and F.Collins, Departments of Internal Medicine and Human Genetics and the Howard Hughes Medical Institute, Univ. of Michigan, Ann Arbor, MI, USA

In order to carry out in vitro and in vivo expression studies on CFTR, it is desirable to have a cDNA clone of the gene which is full length, identical in sequence to the mRNA. Typically, such clones are constructed and propagated in an E. coli host. The full-length clone for CFTR appears to be toxic to E. coli cells, however, so that conventional approaches to cloning are unlikely to work. Using the fragments of the cDNAs10-1, T16-1 and T16-4.5 (Science 245:1066), we have attempted to construct full-length clones, but each trial has yielded primarily grossly rearranged constructs. Only two appeared to be full length and, after sequencing, were also found to have anomalies, one having a 57 bp deletion in exon 6b, the other a single nucleotide insertion in exon 8. To prove that these clones were not merely cloning artifacts, the deleted exon 6b was recloned in parallel with the normal sequence from T16-1. The deletion cloned with an efficiency over 80 fold higher than the normal sequence, and the resultant "normal" clones were also rearranged, whereas the deleted clones were stable. The type and spacing of these clone-stabilizing mutations suggest that at least a portion of CFTR is expressed in the bacterial host and that the unaltered product is toxic to the bacterium. Reducing the quantity of product by switching from a high copy number vector (Bluescript) to a lower copy number vector (~20 copies per cell) of the pBR322 type yields rearranged clones also. Addition of transcriptional terminators at the 5' end to shut off read-through or cryptic transcription from the vector is also ineffective. Likewise, antisense transcription from a lac promoter at the 3' end fails to block the toxic effect of the gene. Numerous other strategies have also been attempted without success. As a partial solution, however, we have been able to create small quantities of the full length clone by ligation of the deletion and insertion clones to each other at a unique AflII restriction site. This method generates sufficient amounts of template for in vitro transcription, the RNA of which is being injected into Xenopus laevis oocytes. Similarly, this approach allows the generation of sufficient cDNA quantities for small-scale transient, complementation assays in which the DNA is administered to the cell by microinjection.